

REMARKS

Applicants had inadvertently renumbered the claims in the response to the second restriction requirement filed on 21 December, 2007. Applicants have herein corrected the error and have numbered the claims to reflect the claim numbers provided in the application as filed. Accordingly, pursuant to the entry of the instant amendment, Claims 1-5, 8-10, 19-23, 25-26, 32-40, 44-47, 56-60, 62-63, and 69-73 are pending; Claims 37-40, 44-47, 56-60, 62-63, and 69-73 are withdrawn; and Claims 6, 7, 11-18, 24, 27-31, 41-43, 48-55, 61, 64-68 and 74-145 are cancelled without prejudice. Claim 1 was amended to recite that the target moiety is a carbohydrate, which was elected pursuant to the second restriction requirement mailed on 11/26/2007. The pending dependent claims have been amended to restore the dependency to the original claim numbers. Withdrawn Claim 37 was amended to correspond to the amendment made in Claim 1, and the remaining withdrawn dependent claims were amended to restore the dependency to the original claim numbers.

I. Amendments to the Specification

In response to the Examiner's objection that the recitation of residue numbers in Claim 10 render the claims indefinite "as it is unclear what these amino acids residues are referred to as no amino acid sequence is presented in the specification either by SEQ ID number or by accession number" (see page 3 of the Office Action), Applicants have amended the specification to address the rejection under 35 USC 112, second paragraph.

II. Sequence Compliance

Applicants have amended the specification to recite the Accession numbers of the *Bacillus amyloliquefaciens* subtilisin being CAA24990 (Appendix 1), and the Accession number of the *Bacillus lenthus* subtilisin (SBL) (Appendix 2) as being P29600. The *Bacillus lenthus* subtilisin is used interchangeably with GG36. For example, the disclosure cites at page 46, lines 17-20 the U.S. patent 5,185,258, which refers to the gene encoding the *Bacillus lenthus* subtilisin (herein referred to as SBL) as GG36 (see for example Examples 2 and 3 of U.S. patent 5,185,258 at column 14, line 51 to column 16, line 47 – Appendix 3), and refers to the product of the *Bacillus lenthus* subtilisins gene product as GG36. For example see column 15, lines 44-46 which refer to the mature

gene product as GG36. Applicants submit that the recitation of the protein accession numbers renders moot the request for a sequence listing.

III. Claim Objections

The Examiner objected to Claims 1-5, 8-10, and 19-36 for containing non-elected subject matter.

Accordingly, applicants have amended Claim 1 to recite "a carbohydrate" as being the target moiety elected in response to the second restriction requirement a mailed on 11/26/2007. The amendment to Claim 1 renders moot the objection of dependent claims 2-5, 8-10, and 19-36.

IV. Claim Rejections under 35 U.S.C. §112, second paragraph

The Examiner has rejected Claim 10 under 35 U.S.C. §112, second paragraph for being "unclear what these amino acid residues are referred to as no amino acid sequence is presented in the specification either by SEQ ID number or by accession number".

As discussed above, Applicants have amended the specification to recite protein accession numbers for the *Bacillus lenthus* subtilisin (SBL) as Protein Accession Number P29600, and Claims 1, 10 and 47 have been amended accordingly to recite the *B. lenthus* subtilisin accession number.

Therefore, the reference residues recited in the claims are no longer undefined.

In light of the foregoing, Applicants respectfully request that the rejection of Claim 10 under 35 U.S.C. §112, second paragraph be withdrawn.

V. Claim Rejections under 35 U.S.C. §112, first paragraph

(A) The Examiner has rejected Claims 1-5, 8-10, 12-29 (now numbered 19-36 as originally filed) under 35 U.S.C. §112, first paragraph as containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention.

In particular, the Examiner states that "these claims are directed to a genus of catalytic antagonists comprising chimeric proteins comprising any subtilisins-type serine hydrolase conjugated through the sulfur group on a cysteine of said hydrolase with a genus of targeting moiety". Further, the Examiner states that "the specification teaches

the structure of only a few such chimeric proteins", and that "the specification fails to describe any other representative species by any identifying characteristics other than the functionality of [the] chimeric protein".

Applicants respectfully traverse the rejection for the following reasons.

First, Applicants have amended Claim 1 to recite that the targeting moiety is a carbohydrate. Therefore, Claim 1 as amended is specifically directed to subtilisin type serine proteases that comprise a carbohydrate targeting moiety, and not to any targeting moiety. Second, in addition to teaching the functionality of the chimeric proteins, the specification also teaches the functionality of the subtilisin serine proteases. For example, at page 13, line 32 to page 14, line 16, the disclosure provides numerous references and teaches that " the "subtilisin type serine proteases" refer to a family of serine hydrolyases based on structural homology to enzymes derived from *Bacillus subtilis*, including subtilisin BPN' (Bott *et al.* (1988) *J. Biol. Chem.* 263: 7895-7906; Siezen and Louise (1997) *Protein Science* 6: 501-523; Bartlett and Rawlings (1994) *Meth. Enzymol.*, 244: 19-61, Academic Press, S.D.). " The specification further teaches the structural characteristics that discern the subtilisin serine protease from the chymotrypsin serine proteases, and discloses that while the amino acid sequences of the subtilisin serine proteases are not entirely homologous, they "exhibit the same or similar type of proteolytic activity". The specification also discloses specific residues of subtilisins type serine proteases that can be substituted and modified to create the claimed catalytic antagonists (for example, see page 42, line 8 to page 43, line 7, and page 44, line 12 to page 46, line 2, and Examples 1-12). Furthermore, the specification also discloses the targeting moieties that can be conjugated to the modified antagonists (for example see page 49, line 17 to page 50, line 20 and Examples 1-12).

Thus, the specification provides adequate description of the claimed catalytic antagonist comprising a targeting moiety that specifically binds to said target molecule said targeting moiety being a carbohydrate attached to an enzyme, said enzyme being a subtilisin-type serine protease" as recited in Claim 1. Applicants submit that the specification sufficiently describes the claimed invention, and shows that Applicants were indeed in possession of the claimed invention at the time the application was filed.

In light of the foregoing, Applicants respectfully request that the rejection of Claims 1-5, 8-10, 12-29 (now numbered 19-36 as originally filed) under 35 U.S.C. §112, first paragraph, be withdrawn.

(B) The Examiner has rejected Claims 1-5, 8-10, 12-29 (now numbered 19-36 as originally filed) under 35 U.S.C. §112, first paragraph, "because the specification, while being enabling for [a] catalytic antagonist comprising [a] chimeric protein of a *Bacillus lenthus* subtilisin-type serine hydrolase conjugated with [a] mannose type oligosaccharide... to target mannose binding lectin...[does] not reasonably provide enablement for any catalytic antagonist comprising [a] chimeric protein comprising any subtilisins type serine hydrolase with a genus of targeting moiety comprising any structural and functional features".

Applicants respectfully traverse the rejection for the following reasons.

As discussed above, Claim 1 has been amended to recite "carbohydrate" as being the targeting moiety. The specification teaches which residues to select for substitution and conjugation, and discloses specific sites in *B. lenthus* subtilisins and other related enzymes (see for example page 42, line 24 to page 43, line 2). Methods for introducing the cysteine residue in the serine hydrolase are described at least at page 44, line 12 to page 46, line 2. The disclosure also teaches methods for coupling the targeting moieties to the cysteine (see page 49, line 16 to page 50, line 20), and provides specific examples relating to the synthesis of carbohydrate modified serine hydrolases (Example 5 at page 81), and the use of a mannosylated serine hydrolase to target mannose binding lectin (Example 6 at page 82).

Therefore, the specification does teach a person skilled in the art to make and use the claimed invention.

In light of the foregoing, Applicants respectfully request that the rejection of Claims 1-5, 8-10, 12-29 (now numbered 19-36 as originally filed) under 35 U.S.C. §112, first paragraph for lacking enablement, be withdrawn.

CONCLUSION

Applicants submit that this paper fully addresses the Office Action mailed March 20, 2008. Applicants believe the pending claims are in condition for allowance and issuance of a formal Notice of Allowance at an early date is respectfully requested. If a telephone conference would expedite prosecution of this application, the Examiner is invited to telephone the undersigned at (650) 846-7636.

This paper is accompanied by a request for Extension of Time under 37 C.F.R. 1.136(a) of THREE months extending the time for response to September 20, 2008. This response is filed prior to the extended deadline and is therefore timely filed. The Commissioner is authorized to charge any fees that may be required in connection with this submission and to credit any overpayments to Deposit Account No. 07-1048 (Attorney Docket No. GC571-2-C1).

Respectfully submitted,

Date: September 19, 2008

/Elena E. Quertermous/

Elena E. Quertermous
Registration No. 47,873

Danisco U.S., Genencor Division
925 Page Mill Road
Palo Alto, CA 94304
Tel: 650-846-7636
Fax: 650-845-6504

Appendix 1

NCBI Entrez Protein My NCBI [Sign In] [Regis

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search for

Display Show

Range: from to Features: CDD

1: CAA24990. Reports precursor subtilis...[gi:773560]

BLink, Conserved Domains, Links

Comment Features Sequence

LOCUS CAA24990 376 aa linear BCT 17-NOV-2004
 DEFINITION precursor subtilisin [Bacillus amyloliquefaciens].
 ACCESSION CAA24990
 VERSION CAA24990.1 GI:773560
 DBSOURCE embl accession X00165.1
 KEYWORDS .
 SOURCE Bacillus amyloliquefaciens
 ORGANISM Bacillus amyloliquefaciens
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1 (residues 1 to 376)
 AUTHORS Wells, J.A., Ferrari, E., Henner, D.J., Estell, D.A. and Chen, E.Y.
 TITLE Cloning, sequencing, and secretion of Bacillus amyloliquefaciens subtilisin in *Bacillus subtilis*
 JOURNAL Nucleic Acids Res. 11 (22), 7911-7925 (1983)
 PUBMED 6316278
 COMMENT On Apr 17, 1995 this sequence version replaced gi:762917.
 Data kindly reviewed (01-JUL-1985) by D.A. Estell.
 FEATURES Location/Qualifiers
 source 1..376
 /organism="Bacillus amyloliquefaciens"
 /db_xref="taxon:1390"
 Protein 1..376
 /product="precursor subtilisin"
 sig_peptide 1..27
 /note="coding sequence signal peptide"
 CDS 1..376
 /coded_by="X00165.1:111..1241"
 /note="tgg start"
 /transl_table=11
 /db_xref="GOA:P00782"
 /db_xref="PDB:1A2Q"
 /db_xref="PDB:1AK9"
 /db_xref="PDB:1AQN"
 /db_xref="PDB:1AU9"
 /db_xref="PDB:1DU1"
 /db_xref="PDB:1GNS"
 /db_xref="PDB:1GNV"
 /db_xref="PDB:1LW6"
 /db_xref="PDB:1S01"
 /db_xref="PDB:1S02"
 /db_xref="PDB:1SBH"
 /db_xref="PDB:1SBI"
 /db_xref="PDB:1SBN"
 /db_xref="PDB:1SBT"

```
/db_xref="PDB:1SIB"
/db_xref="PDB:1SPB"
/db_xref="PDB:1ST2"
/db_xref="PDB:1SUA"
/db_xref="PDB:1SUB"
/db_xref="PDB:1SUC"
/db_xref="PDB:1SUD"
/db_xref="PDB:1SUE"
/db_xref="PDB:1SUE"
/db_xref="PDB:1TM1"
/db_xref="PDB:1TM3"
/db_xref="PDB:1TM4"
/db_xref="PDB:1TM5"
/db_xref="PDB:1TM7"
/db_xref="PDB:1TMG"
/db_xref="PDB:1TO1"
/db_xref="PDB:1TO2"
/db_xref="PDB:1UBN"
/db_xref="PDB:1VS1"
/db_xref="PDB:1YJA"
/db_xref="PDB:1YJB"
/db_xref="PDB:1YJC"
/db_xref="PDB:2SBT"
/db_xref="PDB:2SIC"
/db_xref="PDB:2SNI"
/db_xref="PDB:2ST1"
/db_xref="PDB:3SIC"
/db_xref="PDB:5SIC"
/db_xref="UniProtKB/Swiss-Prot:P00782"
```

ORIGIN

```
1 misllfalal iftmagsts saaqagksng ekkyivgfkq tmstmsaakk kdvisekggk
61 vqkqfkvyda asatlinekav kelkkdpsva yveedvhava yaqsvpygvs qikapalhsq
121 gytgsvnvkaa vidsgidssh pdlkvaggas mvpsetnpfq dnnshgthva gtvaaalnnsi
181 gvlgvapsas lyavkvlgad gsggyswiin giewaiannm dvimmslgrp sgsaalkaav
241 dkavasgvvv vaaagnegts gssttvgypg kypsviavga vdssnqrasf ssvgpeldvm
301 apgvsigstl pgnykgayng tmasphvag aaalilskhp nwtnqvrsst lentttkldg
361 sfyygklin vqaaaq
```

//

Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

Last update: Thu, 03 Jul 2008 Rev. 132917

Appendix 2

NCBI    **Protein** [Sign In] [Regis]

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search **Protein** for

Display **GenPept** Show 5

Range: from to Features: CDD

P1: P29600. Reports Subtilisin Savinase...[gi:267048]

BLINK, Conserved Domains, Links

Comment Features Sequence

LOCUS P29600 269 aa linear BCT 15-JAN-2008
 DEFINITION Subtilisin Savinase (Alkaline protease).
 ACCESSION E29600
 VERSION P29600.1 GI:267048
 DBSOURCE swissprot: locus SUBS_BACLE, accession P29600;
 class: standard.
 created: Apr 1, 1993.
 sequence updated: Apr 1, 1993.
 annotation updated: Jan 15, 2008.
 xrefs: 1C9J, 1C9M, 1C9N, 1GCI, 1IAV, 1JEA, 1NDQ, 1NDU, 1Q5P, 1SVN,
 1TK2
 xrefs (non-sequence databases): PDBsum:1C9J, PDBsum:1C9M,
 PDBsum:1C9N, PDBsum:1GCI, PDBsum:1IAV, PDBsum:1JEA, PDBsum:1NDQ,
 PDBsum:1NDU, PDBsum:1Q5P, PDBsum:1SVN, PDBsum:1TK2, LinkHub:P29600,
 GO:0004291, InterPro:IPR000209, InterPro:IPR015500,
 Gene3D:G3DSA:3.40.50.200, PANTHER:PTHR10795, Pfam:PF00082,
 PRINTS:PR00723, PROSITE:PS00136, PROSITE:PS00137, PROSITE:PS00138
 KEYWORDS 3D-structure; Calcium; Hydrolase; Metal-binding; Protease;
 Secreted; Serine protease; Sporulation.
 SOURCE Bacillus lentus
 ORGANISM Bacillus lentus
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1 (residues 1 to 269)
 AUTHORS Betzel,C., Klupsch,S., Papendorf,G., Hastrup,S., Branner,S. and
 Wilson,K.S.
 TITLE Crystal structure of the alkaline proteinase Savinase from Bacillus
 lentus at 1.4 Å resolution
 JOURNAL J. Mol. Biol. 223 (2), 427-445 (1992)
 PUBMED 1738156
 REMARK X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS).
 REFERENCE 2 (residues 1 to 269)
 AUTHORS Remerowski,M.L., Pepermans,H.A., Hilbers,C.W. and Van De Ven,F.J.
 TITLE Backbone dynamics of the 269-residue protease Savinase determined
 from ¹⁵N-NMR relaxation measurements
 JOURNAL Eur. J. Biochem. 235 (3), 629-640 (1996)
 PUBMED 8654411
 REMARK STRUCTURE BY NMR.
 REFERENCE 3 (residues 1 to 269)
 AUTHORS Kuhn,P., Knapp,M., Soltis,S.M., Ganshaw,G., Thoene,M. and Bott,R.
 TITLE The 0.78 Å structure of a serine protease: Bacillus lentus
 subtilisin
 JOURNAL Biochemistry 37 (39), 13446-13452 (1998)
 PUBMED 9753430
 REMARK X-RAY CRYSTALLOGRAPHY (0.78 ANGSTROMS).

COMMENT [FUNCTION] Subtilisin is an extracellular alkaline serine protease, it catalyzes the hydrolysis of proteins and peptide amides.
 [CATALYTIC ACTIVITY] Hydrolysis of proteins with broad specificity for peptide bonds, and a preference for a large uncharged residue in P1. Hydrolyzes peptide amides.
 [COFACTOR] Binds 2 calcium ions per subunit.
 [SUBCELLULAR LOCATION] Secreted.
 [BIOTECHNOLOGY] Used as a detergent protease. Sold under the name Savinase by Novozymes.
 [MISCELLANEOUS] Secretion of subtilisin is associated with onset of sporulation, and many mutations which block sporulation at early stages affect expression levels of subtilisin. However, subtilisin is not necessary for normal sporulation.
 [SIMILARITY] Belongs to the peptidase S8 family.

FEATURES Location/Qualifiers

source	1..269 /organism="Bacillus lenthus" /db_xref="taxon:1467"
Protein	1..269 /product="Subtilisin Savinase" /EC_number="3.4.21.62"
Region	1..269 /region_name="Mature chain" /experiment="experimental evidence, no additional details recorded" /note="Subtilisin Savinase. /FTId=PRO_0000076418."
Site	2 /site_type="metal-binding" /experiment="experimental evidence, no additional details recorded" /note="Calcium 1."
Region	6..10 /region_name="Helical region" /experiment="experimental evidence, no additional details recorded"
Region	13..18 /region_name="Helical region" /experiment="experimental evidence, no additional details recorded"
Region	27..33 /region_name="Beta-strand region" /experiment="experimental evidence, no additional details recorded"
Site	32 /site_type="active" /experiment="experimental evidence, no additional details recorded" /note="Charge relay system."
Site	40 /site_type="metal-binding" /experiment="experimental evidence, no additional details recorded" /note="Calcium 1."
Region	43..48 /region_name="Beta-strand region" /experiment="experimental evidence, no additional details recorded"
Region	59..61 /region_name="Beta-strand region" /experiment="experimental evidence, no additional details recorded"

```
recorded"
Region 62.71
/region_name="Helical region"
/experiment="experimental evidence, no additional details
recorded"
Site 62
/site_type="active"
/experiment="experimental evidence, no additional details
recorded"
/note="Charge relay system."
Site 73
/site_type="metal-binding"
/experiment="experimental evidence, no additional details
recorded"
/note="Calcium 1 (via carbonyl oxygen)."
Region 75.78
/region_name="Beta-strand region"
/experiment="experimental evidence, no additional details
recorded"
Site 75
/site_type="metal-binding"
/experiment="experimental evidence, no additional details
recorded"
/note="Calcium 1."
Site 77
/site_type="metal-binding"
/experiment="experimental evidence, no additional details
recorded"
/note="Calcium 1 (via carbonyl oxygen)."
Site 79
/site_type="metal-binding"
/experiment="experimental evidence, no additional details
recorded"
/note="Calcium 1 (via carbonyl oxygen)."
Region 87..92
/region_name="Beta-strand region"
/experiment="experimental evidence, no additional details
recorded"
/note="Calcium 1 (via carbonyl oxygen)."
Region 102..114
/region_name="Helical region"
/experiment="experimental evidence, no additional details
recorded"
Region 118..122
/region_name="Beta-strand region"
/experiment="experimental evidence, no additional details
recorded"
Region 126..128
/region_name="Beta-strand region"
/experiment="experimental evidence, no additional details
recorded"
Region 131..142
/region_name="Helical region"
/experiment="experimental evidence, no additional details
recorded"
Region 146..150
/region_name="Beta-strand region"
/experiment="experimental evidence, no additional details
recorded"
Region 162..164
/region_name="Hydrogen bonded turn"
```

/experiment="experimental evidence, no additional details recorded"
Site 163
/site_type="metal-binding"
/experiment="experimental evidence, no additional details recorded"
/note="Calcium 2 (via carbonyl oxygen)."
Site 165
/site_type="metal-binding"
/experiment="experimental evidence, no additional details recorded"
/note="Calcium 2 (via carbonyl oxygen)."
Region 168..174
/region_name="Beta-strand region"
/experiment="experimental evidence, no additional details recorded"
Site 168
/site_type="metal-binding"
/experiment="experimental evidence, no additional details recorded"
/note="Calcium 2 (via carbonyl oxygen)."
Region 190..195
/region_name="Beta-strand region"
/experiment="experimental evidence, no additional details recorded"
Region 197..203
/region_name="Beta-strand region"
/experiment="experimental evidence, no additional details recorded"
Region 204..206
/region_name="Hydrogen bonded turn"
/experiment="experimental evidence, no additional details recorded"
Region 207..211
/region_name="Beta-strand region"
/experiment="experimental evidence, no additional details recorded"
Region 214..231
/region_name="Helical region"
/experiment="experimental evidence, no additional details recorded"
Site 215
/site_type="active"
/experiment="experimental evidence, no additional details recorded"
/note="Charge relay system."
Region 237..246
/region_name="Helical region"
/experiment="experimental evidence, no additional details recorded"
Region 254..257
/region_name="Helical region"
/experiment="experimental evidence, no additional details recorded"
Region 264..267
/region_name="Helical region"
/experiment="experimental evidence, no additional details recorded"
ORIGIN
1 aqsvpwglsr vqapaahnrg ltgsgvkav ltdgisthpd lnirggasfv pgepstqdgn

61 ghgthvagtiaalnnnsigvl gvapsaelyavkvlgasgsg svssiaqgle wagnngmhva
121 nlslgspspstaleqavnsatsrgvlvaa sgnsgagsis yparyanama vgatdqnnnr
181 asfsqyagl divapgvnvq stypgstyas lngttsmatph vagaaalvkvknpswsnvqi
241 rnhlkntats lgstnlygsg lvnaeaatr

//

[Disclaimer](#) | [Write to the Help Desk](#)
[NCBI](#) | [NLM](#) | [NIH](#)

Last update: Thu, 03 Jul 2008 Rev. 132917

Appendix 3



US007413877B2

(12) **United States Patent**
Collier et al.

(10) **Patent No.:** **US 7,413,877 B2**
(b4) **Date of Patent:** ***Aug. 19, 2008**

(54) **BACTERIAL EXPRESSION OF
BOWMAN-BIRK PROTEASE INHIBITORS
AND VARIANTS THEREOF**

WO WO 00/05406 2/2000

(75) Inventors: **Katherine Collier**, Los Altos, CA (US); **Grant Ganshaw**, Tracy, CA (US); **Hans De Nobel**, Almere (NL); **Scott D. Power**, San Bruno, CA (US); **Antita Van Kimmenade**, San Bruno, CA (US); **Mark Kolkman**, Oegstgeest (NL); **Jeffrey Miller**, Santa Cruz, CA (US); **Brian Schmidt**, Half Moon Bay, CA (US); **Gudrun Vogtentanz**, Santa Clara, CA (US); **David Estell**, San Mateo, CA (US)

(73) Assignee: **Genencor International, Inc.**, Palo Alto, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 185 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **10/984,514**

(22) Filed: **Nov. 8, 2004**

(65) **Prior Publication Data**

US 2005/0202535 A1 Sep. 15, 2005

Related U.S. Application Data

(60) Provisional application No. 60/531,207, filed on Dec. 19, 2003; provisional application No. 60/531,189, filed on Dec. 19, 2003; provisional application No. 60/530,954, filed on Dec. 19, 2003; provisional application No. 60/520,403, filed on Nov. 13, 2003; provisional application No. 60/518,154, filed on Nov. 6, 2003.

(51) **Int. Cl.**
C12P 29/00 (2006.01)
C07H 21/04 (2006.01)
C12N 1/21 (2006.01)

(52) **U.S. Cl.** **435/71.1; 435/252.3; 435/252.31;**
536/23.1

(58) **Field of Classification Search** None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,411,873 A	5/1995	Adams et al.	435/69.1
5,429,950 A	7/1995	Power et al.	435/252.31
5,679,543 A	10/1997	Lawrie	435/69.1
6,063,611 A	5/2000	Van Solingen	435/209
6,537,968 B1	3/2003	Lendley et al.	514/42
6,872,563 B1 *	3/2005	Beckwith et al.	435/252.3

FOREIGN PATENT DOCUMENTS

WO WO 88/02025 * 8/1987

OTHER PUBLICATIONS

Flecker "Chemical synthesis, molecular cloning and expression of gene coding for a Bowman-Birk-type protease inhibitor," *Eur. J. Biochem.*, 1987, 166, 151-4.

Büllings et al., "A Growth-regulated Protease Activity that is Inhibited by the Anticarcinogenic Bowman-Birk Protease Inhibitor," *Proc. Natl. Acad. Sci. USA*, 89:3120-3124 (1992).

Birk, "The Bowman-Birk Inhibitor: Trypsin- and Chymotrypsin-inhibitor from Soybeans," *Int. J. Peptide Protein Res.*, 25:113-131 (1985).

Bode et al., "Natural Protein Proteinase Inhibitors and Their Interaction with Proteinases," *Eur. J. Biochem.*, 204:433-451 (1992).

Chen et al., "Reactive Sites of an Anticarcinogenic Bowman-Birk Proteinase Inhibitor are Similar to Other Trypsin Inhibitors," *The Journal of Biological Chemistry*, 267(3):1990-1994 (1992).

Christmann et al., "The Cystine Knot of a Sausage-type Protease Inhibitor as a Structural Scaffold for *Escherichia coli* Cell Surface Display of Conformationally Constrained Peptides," *Protein Engineering*, 12(9):797-806 (1999).

Fernari et al., "Transcription of *Bacillus subtilis* Subtilisin and Expression of Subtilisin in Sporulation Mutants," *Journal of Bacteriology*, 170(1):289-295 (1988).

Flecker et al., "Chemical Synthesis, Molecular Cloning and Expression of Gene Coding for a Bowman-Birk-type Proteinase Inhibitor," *Eur. J. Biochem.*, 166:151-156 (1987).

Iahn et al., "Regulatory Inputs for the Synthesis of ComX, the Competence Transcription Factor of *Bacillus subtilis*," *Molecular Microbiology*, 21(4):763-775 (1996).

Henner et al., "Location of the Targets of the *hpr-97*, *sacU32(Hy)*, and *sacC36(Hy)* Mutations in Upstream Regions of the Subtilisin Promoter," *J. Bact.*, 170(1):296-300 (1988).

Hengen "Purification of His-Tag Fusion Proteins from *Escherichia coli*," *TIBS*, 20:285-286 (1995).

Kajino et al., "A Protein Disulfide Isomerase Gene Fusion Expression System that Increases the Extracellular Productivity of *Bacillus brevis*," *Applied and Environmental Microbiology*, 66(2):638-642 (2000).

Kemperman et al., "Identification and Characterization of Two Novel Clostridial Bacteriocins, Circularin A and Closticin 574," *Applied and Environmental Microbiology*, 69(3):1589-1597 (2003).

Kennedy "The Bowman-Birk Inhibitor from Soybeans as an Anticarcinogenic Agent," *Am. J. Clin. Nutr.*, 68:1406S-12S (1998).

Landon, *Methods in Enzymology*, Cleavage at Aspartyl-Prolyl Bonds, Academic Press, Inc., pp. 145-149 (1977).

Lidell et al., "An Autocatalytic Cleavage in the C Terminus of the Human MUC2 Mucin Occurs at the Low pH of the Late Secretory Pathway," *The Journal of Biological Chemistry*, 278(16):13944-13951 (2003).

(Continued)

Primary Examiner—Cecilia Tsang

Assistant Examiner—Christina Marchetti Brad

(57) **ABSTRACT**

The present invention provides compositions and methods related to the expression of Bowman-Birk protease inhibitors and variants thereof in bacterial species. The present invention further provides fusion nucleic acids, vectors, fusion polypeptides, and processes for obtaining the Bowman-Birk protease inhibitors.

OTHER PUBLICATIONS

Lin et al., "The 0.25-nm X-ray Structure of the Bowman-Birk-type Inhibitor from Mung Bean in Ternary Complex with Porcine Trypsin," *Eur. J. Biochem.*, 212:549-555 (1993).

Liu, *Soybeans, Chemistry, Technology and Utilization*, pp. 32-35, Aspen Publishers, Inc., Gaithersburg, Maryland (1999).

Meima et al., "The *isbhDC* Operon of *Bacillus subtilis* Encodes Thiol-Oxidoreductases Required for Competence Development," *The Journal of Biological Chemistry*, 277(19):6994-7001 (2002).

Neidhardt et al., "Culture Medium for Enterobacteria," *Journal of Bacteriology*, 119(3):736-747 (1974).

Odani et al., "Studies on Soybean Trypsin Inhibitors. IV. Complete Amino Acid Sequence and the Anti-proteinase Sites of Bowman-Birk Soybean Proteinase Inhibitor," *J. Biochem.*, 71:839-848 (1972).

Paine et al., "An Alternative Approach to Depigmentation by Soybean Extracts via Inhibition of the PAR-2 Pathway," *J. Invest. Dermatol.*, 116:587-595 (2001).

Perego, "Integration Vectors for Genetic Manipulation in *Bacillus subtilis*," *Bacillus subtilis* and Other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics, Sonenshein, Hoch and Losick (eds.), *American Society for Microbiology*, Washington D.C., pp. 615-624 (1993).

Sahu et al., "Inhibition of Human Complement by a C3-Binding Peptide Isolated from a Phage-Displayed Random Peptide Library," *The Journal of Immunology*, 157:884-891 (1996).

Seeboth et al., "In-vitro Cleavage of a Fusion Protein Bound to Cellulose Using the Soluble *yscF*s (Kex₂) Variant," *Appl. Microbiol. Biotechnol.*, 37:621-625 (1992).

Sigalas et al., "A Particularly Labile Asp-Pro Bond in the Green Mamba Muscarinic Toxin MTX2 Effect of Protein Conformation on the Rate of Cleavage," *FEBS Letters*, 371:171-175 (1995).

Shaw et al., "A Novel Combination of Two Classic Catalytic Schemes," *J. Mol. Biol.*, 320:303-309 (2002).

Song et al., "Kunitz-type Soybean Trypsin Inhibitor Revisited: Refined Structure of its Complex with Porcine Trypsin Reveals an Insight into the Interaction Between a Homologous Inhibitor from *Erythrina caffra* and Tissue-type Plasminogen Activator," *J. Mol. Biol.*, 275:347-363 (1998).

Van Tilburgh et al., "Fluorogenic and Chromogenic Glycosidases as Substrates and Ligands of Carbohydrases," *Methods in Enzymology*, Academic Press, Inc., pp. 45-59 (1988).

Voss et al., "Crystal Structure of the Bifunctional Soybean Bowman-Birk Inhibitor at 0.28-nm Resolution," *Eur. J. Biochem.*, 242:122-131 (1996).

Werner et al., "Three-Dimensional Structure of Soybean Trypsin/Chymotrypsin Bowman-Birk Inhibitor in Solution," *Biochemistry*, 31:999-1010 (1992).

Wolfson et al., "Modularity of Protein Function I: Chimeric Interleukin 1 β Containing Specific Protease Inhibitor Loops Retain Function of Both Molecules," *Biochemistry*, 32:5327-5331 (1993).

* cited by examiner

FIGURE 1.

aprE promoter region

EcoRI

1 AATTCTCCAT TTTCTCTGC TATCAAAATA ACAGACTCGT GATTTCCAA

aprE promoter region

51 ACAGAGCTTC AAAAAGGCCT CTGCCCTTG CAAATCGGAT GCCTGTCTAT

aprE promoter region

EagI

NotI

101 AAAATTCCCG ATATTGGTTA AACAGCGCG CAATGGCGC CGCATCTGAT

aprE promoter region

151 GTCTTGCCTT GGCGAATGTT CATCTTATTCTT CTTCCCTCCCT CTCATAAATT

aprE promoter region

201 TTTCATCTT ATCCCTTTTC TGTAAGTTT ATTTTCAGA ATACTTTAT

aprE promoter region

251 CATCATGCTT TGAAAAAAATA TCACGATAAT ATCCATTGTT CTCACCGAAG

aprE promoter region

301 CACACGCAGG TCATTGAAAC GAATTTTTTC GACAGGAATT TGCCGGGACT

aprE promoter region

351 CAGGAGCATT TAACCTAAAAA AAGCATGACA TTTCAGCATA ATGAACATTT

aprE promoter region

401 ACTCATGTCT ATTTCTGTTTCTTCTGTAT GAAAATAGTT ATTTCGAGTC

aprE promoter region

451 TCTACGGAAA TAGCGAGAGA TGATATACTT AAATAGAGAT AAAATCATCT

aprE promoter region

501 CAAAAAAAATG GGTCTACTAA AATATTATTC CATCTATTAC AATAAATTCA

aprE promoter region

551 CAGAAATAGTC TTTCAGTAA GTCTACTCTG AATTTTTTTA AAAGGAGAGG

AprE signal peptide

aprE promoter region

601 M R S K K L W I S L L F A L T T
GTAAGAGTG AGAACCAAAA AATGTTGGAT CAGCTTGTG TTTCCGTAA

FIGURE 1. (cont.)

AprE signal peptide

BCE103

· L I F T M A F S N M S A Q A D D
651 CGTTAATCTT TAGCATGGCG TTCAGCAACA TGCTCGCGA GGCTGATGAT

BCE103

Y S V V E E H G Q L S I S N G E L .
701 TATTCAGTT TAGAGGAACA TGGGCAACTA AGTATTAGTA ACGGTGAATT

BCE103

NcoI

· V N E R G E Q V Q L K G M S S H G .
751 AGTCAATGAA CGAGGCAAC AGATTCAGTT AAAAGGGATG AGTTCCATG

BCE103

NcoI

· L Q W Y G Q F V N Y E S M K W L
801 GTTGCAATG GTACGGTCAA TTTGTAACAT ATGAAAGCAT GAAATGGCTA

BCE103

R D D W G I T V F P R A A M Y T S S .
851 AGAGATGATT GGGGAATAAC TGATTTCCGA GCAGCAATGT ATACCTCTTC

BCE103

· G G Y I D D P S V K E K V K E T V .
901 AGGAGGATAT ATTGACGATC CATCAGTAA GGAAAAAGTA AAAGAGACTG

BCE103

· E A A I D L G I Y V I I D W H I .
951 TTGAGGCTGC GATAGACCTT CACATATATG TGATCATGAA TTGGCATATC

BCE103

L S D N D P N I Y K E E A K D F F .
1001 CTTTCAGACA ATGACCGAA TATATATAAA GAAGAAGCGA AGGATTTCTT

BCE103

· D E M S E L Y G D Y P N V I Y E I .
1051 TGATGAAATG TCAGAGTTGT ATGGAGACTA TCCGAATGTG ATATACGAA

BCE103

· A N E P N G S D V T W D N Q I K .
1101 TTGCAAATGA ACCGAATGGT AGTGTGTTA CGTGGACAA TCAAAATAAA

BCE103

P Y A E E V I P V I R D N D P N N .
1151 CCGTATGCG AGAAAGTGT CTCGGTTATT CGTGACAATG ACCCTAATAA

FIGURE 1. (cont.)

BCE103

 . I V I V G T G T W S Q D V H H A A .
 1201 CATTGTTATT GTAGGTACAG GTACATGGAG TCAGGATGTC CATCATGCAG

BCE103

 . D N Q L A D P N V M Y A F H F Y .
 1251 CCGATAATCA GCTTGCAGAT CCTAACGTCA TGATGCAATT TCATTTTAT

BCE103

 . A G T H G Q N L R D Q V D Y A L D .
 1301 CGACGGAACAC ATGGACAAAAA TTACGAGAC CAAGTAGATT ATGCATTAGA

BCE103

 . Q G A A I F V S E W G T S A A T G .
 1351 TCAAGGACCA GCGATATTG TTATGATG GGGGACAACT GCAGCTACAG

BCE103

 . D G G V F L D E A Q V W I D F M .
 1401 GTGATGGTGG TGTGTTTTA GATGACAC AAAGTGTGGAT TGACTTTATG

BCE103

 . D E R N L S W A N N S L T H K D E .
 1451 GATGAAAGAA ATTTAACGTC GGCCAACTGG TCTCTAACGC ATAAGGATGA

BCE103

 PstI

 . S S A A L M P G A N P T G G W T E .
 1501 GTCATCTCA GCGTTAACATGC CAGGTGCAA TCCAACTGGT GGTTGGACAG

BCE103

 . A E L S P S G T F V R E K I R E .
 1551 AGGCTGAACCT ATCTCCATCT GGATCACATTG TGAGGGAAAA AATAAGAGAA

BCE103

 1st CBD Linker

 S A S I P P S D P T P P S D P G E .
 1601 TCAGCATCTA TTCCGCCAAG CGATCCAACA CGGCCATCTG ATCCAGGAGA

BBI

 fusion site

 1st CBD Linker

 BamHI SacI

 . P D P D D E S S K P C C D Q C A C .
 1651 ACCGGATCCA GACGATGAGA GCTCTAAACC CTGTTGCGAT CAATGGCGAT

FIGURE 1. (cont.)**BBI**

· T K S N P P Q C R C S D M R L N
1701 GTACGAAATC AAATCCTCCA CAGTGTGGT GTTCCGATAT GCGTCTGAAT

BBI

S C H S A C K S C I C A L S Y P A ·
1751 AGCTGTCATA GTGCATGCAA AAGCTGTATC TGCGCCCTGA GTTATCCAGC

BBI

Sali

· Q C F C V D I T D F C Y E P C K P ·
1801 TCAATGTTTT TGCCTCGACA TCACGGACTT CTGCTATGAG CCATGTAAC

6xHIS

BBI

· S E D D K E N H H H H H H Stop (SEQ ID NO:2)
1851 CAAGCGAGGA CGATAAAGAG ACCATCATC ACCATCACCA TAAAAGTTA

LAT terminator

HindIII

1901 ACAGAGGACG GATTCCTGA AGGAAATCCG TTTTTTATT TTTAAGCTTG (SEQ ID NO:1)

FIGURE 2.

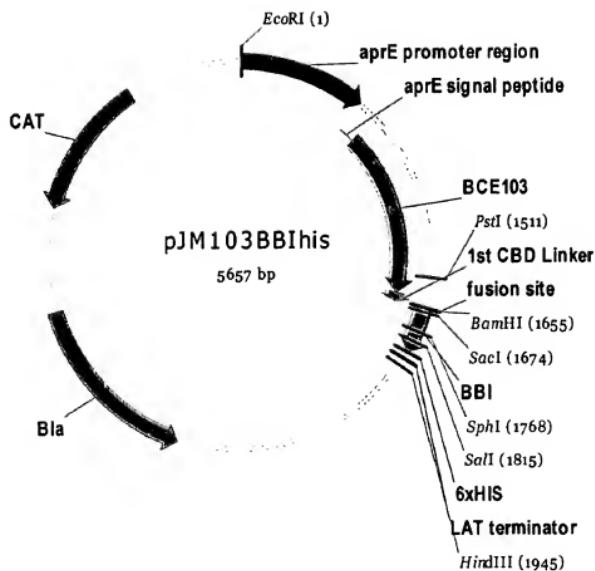


FIGURE 3.

```

----- 12BBIck81 -----
BamHI          SacI
----- 12BBIck81 -----
D P D D E S S K P C C D Q C A C Y
1 GGATCCAGAC GATGAGAGCT CTAAACCCCTG TTGGGATCAA TGCGCATGTT
CCTAGGTCTG CTACTCTCGA GATTGGGAC AACGCTAGTT ACGCGTACAA

----- 12BBIck81 -----
----- PstI          EcoRI -----
N L Y G W T C R C S D M R L N S
51 ATAATTGTA TGGGTGACT TGTGCGTGA GCGATATGCG TCTGAATTCC
TATTAACAT ACCCACCTGA ACAGCGACGT CGCTATACGC AGACTTAAGG

----- 12BBIck81 -----
----- C H S A C K S C A C Y N L Y G W T -----
101 TGTGATAGT CCTGCAAAAG CTGCGCATGT TATAACCTGT ACGGGTGGAC
ACAGTATCAC GGACGTTTC GACCGTACA ATATGGACA TGCCCACCTG

----- 12BBIck81 -----
----- SalI -----
----- C F C V D I T D F C Y E P C K P S -----
151 CTGTTTTG GTCGACATCA CGGACTTCTG CTATGAGCCA TGTAACCAA
GACAAAAACG CAGCTGTAGT GCCTGAAGAC GATACTCGT ACATTTGGTT

----- 12BBIck81 -----
----- E D D K E N * (SEQ ID NO:4) -----
201 GCGAGGAGCAC TAAAGAGAAC TAA (SEQ ID NO:3)
CGCTCCCTGCT ATTTCTCTTG ATT

```

FIGURE 4.

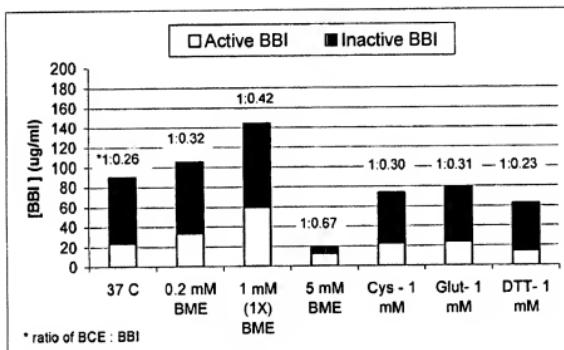


FIGURE 5.

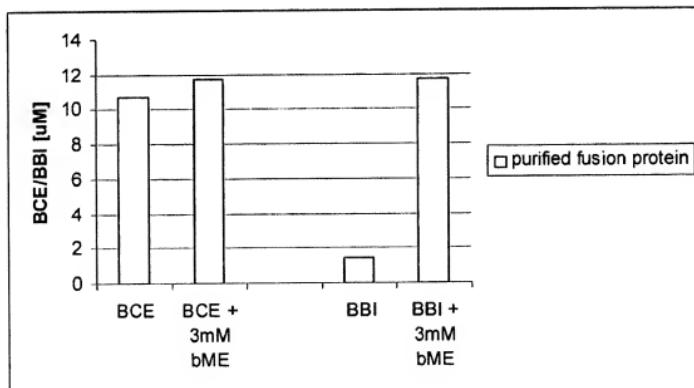


FIGURE 6.

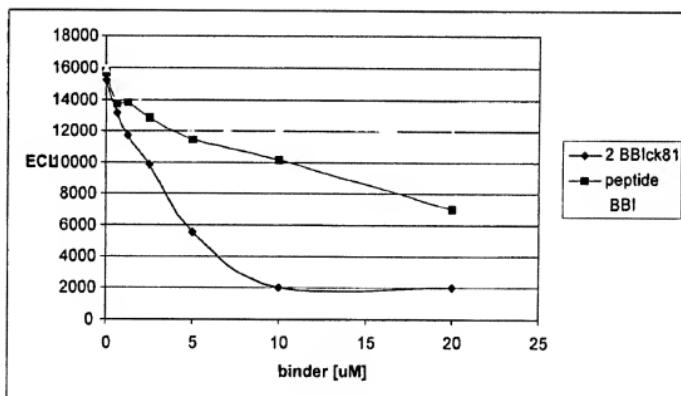


FIGURE 7.

hipDI

AprE signal cleavage site

BshHII NheI BsrGI

S A Q A S D V V Q L K K D T F D D ·

1 AGCCGCAGG CTAGCGATGT TGTCACACTG AAAAAAGACA CTTTGAGCA
TCGGCGCTCC GATCGCTACA ACATGTGAC TTTTTCTGT GAAAGCTGCT

hipDI

. F I K T N D L V L A E F F A P W C ·

51 CTTCATCAA ACAATGACCTTGTTCTTG TGAACTTTCG CGCCCGTGTG
GAACTAGTTTGTTTACTGAAACAAGAAGC ACTTTAAAG CGGGGACCA

hipDI

. G H C K A L A P E Y E E A A T T

101 CGCGTCAGTCAAGCTT GCTCTGAGT ACAGGAAGC TGCAACTACA
CGCCAGTGCAGTTGAGAA CGAGACTCA TGCTCTTCG ACCTTGATGT

hipDI

. L K E K N I K L A K V D C T E E T ·

151 CTGAAAGAAA AGAACATCAA ACTTGCTAAA GTAGACTGCA CAGAAGAC
GACTTCTTTCTTGAGTT TGAAACGATT CATCTGACGT GTCTCTCTG

hipDI

. D L C Q Q H G V E G Y P T L K V F ·

201 TGATCTTGGC CAACACATG GTGTTGAGGG CTACCCAACT CTAAAGCTTT
ACTAGAAACG GTTGTGAC CACAACTCCC GATGGGTGA GAATTTCAAA

hipDI

. R G L D N V S P Y K G Q R K A A

251 TCGTGGCT TGACAACTGA TCTCCCTACA AAGGTCAAGC TAAAGCTGCT
AGCCACCGGA ACTGTTGAT AGAGGAATGT TTCCAGTTCG ATTCGACGA

hipDI

. A I T S Y M I K Q S L P A V S E V ·

301 GCAATCACTT CATACTATGAT CAAACATCT CTGCTGCTG TATCTGAAGT
CGTTAGTGAATGTTAGA GACGGACGAC ATAGACTTC

hipDI

. T K D N L E E F K K A D K A V L V ·

351 TACAAAAGAC AACCTGAAAG AATTTAAAAA AGCTGACAAA CCTGTTCTTG
ATGTTTCTG TTGGAACATTC TTAAATTTT TCGACTGTT CGACAAAGAC

hipDI

. A Y V D A S D K A S S E V P T Q ·

401 TTGCTTATGAT AGATGCTCT GACAAAGCAT CTAGCGAAAGT TTCACTCAA
AACGAATACA TCTAGGAGA CTGTTCTGA GATCGCTCA AAATGAGTT

hipDI

. V A E K L R D N Y P F G S S S D A ·

451 GTTGTGAAAG AACCTGGCGA TAATACCCA TTGGCTCTA GCTCTGATGC
CAACGACTT TTGACGGCTT ATTGATGGT AAGCGAGAT CGGAGACTACG

FIGURE 7 (cont.)

hiPDI

 501 . A L A E A E G V K A P A I V L Y K .
 TGCAGCTGGCT GAAGCTGAGG GCGTAAAGAC ACCTGCTATT GTTCTTTACA
 ACCTGACCGA CCTCGACTCC CGCAATTTCG TGGACGATAA CAAGAAATGT

hiPDI

 551 . D F D E G K A V F S E K F E V E .
 AAAGACTTGA TGAAAGTAAA GGGTTTTCT CTGAAAATT CGAAGTAGAG
 TTCTGAAACT ACTTCCATTG CGCCAAAAGA GACTTTTAA GCTTCATCTC

hiPDI

 601 . A I E K F A K T G A T P L I G E I .
 GCAATCGAAA AATTCGCTAA AACAGCTGTCT ACCTCCACTTA TTGGCGAAAT
 CCTTAGCTTT TTAAAGCATT TTGTCACAGA TGAGGTGAAT AACCGCTTTA

hiPDI

 651 . G P E T Y S D Y M S A G I P L A Y .
 CGGACCTGAA ACTTACTCTG ATTACATGTC AGCTGGCATC CCTCTGGCAT
 CGCTTGGACTT TGAATGAGAC TAATGTACAG TCGACCGTAG GGAGACCGTA

SapI

 701 . I F A E T A E E R K E L S D K L .
 ACATTTTCCG TGAAACAGT GAAGAGCGTA AAGACTCTAG CGACAAACTT
 TTGAAAAGGG ACTTTCGGA CCTCTCGCAT TTCTTGAGTC GCTGTTGAA

hiPDI

 751 . K P I A B A Q R G V I N F G T I D .
 AAACCAATCG CTGAACAGCT ACCTGGGGTT ATTAACTTTCG TTACTATTTGA
 TTGCGTTAGC GACTTCGAGT TGACCCCGAA TAATTGAAAC CATGATAACT

hiPDI

 801 . A K A F G A H A G N L N L K T D K .
 CGCTAAAGCA TTGGTGTCTC AGCTGGAAA CCTGAATCTG AAAACTGACA
 CGGATTTCTG AAACCAAGAG TGGCACCTTT GGACTTACAG TTTTGACTGT

hiPDI

 851 . F P A F A I Q E V A K N Q K F P .
 AATTCCTGC TTGCGAACAT CGAAGAGTTG CTAAAAAACCA AAAATTCCTT
 TTGAGGACG AAAGCGTTG TTCTTCAC GATTTTTGGT TTTTAAGGGA

hiPDI

 901 . F D Q E K E I T F E A I K A F V D .
 TTGATCAAG AAAAGAAAT TACTTTGAA CGGATCAAG CATTCGTGAA
 AAACAGTTC TTTCCTTAA ATGAAAACCTT CGCTAGTTTC GAAAGCAACT

hiPDI

 951 . D F V A G K I E P S I K S E P I P .
 CGATTTGTTG GCTGGTAAA TCGAACCGAG CATCAATCA GAAACCAATCC
 GCTAAACAA CGACCAATTG AGCTGGTTC GTAGTTAGT CTGGTTAGG

hiPDI

 1001 . E K Q E G G P V T V V V A K N Y N .
 CTGAAACAA AGAAGGCTCT GATACGTGAG TTGTAGCTAA AACATCAAT
 GACTTTTGTG TCTTCAGGAGA CAATGACATC AACATCGATT TTGATGTTA

FIGURE 7. (cont.)

hipDI

```

-----  

E I V L D D T K D V L I E F Y A P -  

1051 GAAATCGTTC TGGACGATAC TAAAGATGA TAAATGGAT TTTACGCTCC  

CTTTAGCAAG ACCCTGATAG ATTCTACAT ATTTAACCTA AAATGCGAGG

```

hipDI

```

-----  

W C G H C K A L A P K Y E E L G A -  

1101 TTGGTGGTGT CACTGCAAG CTCTTGCTCC TAAATACCAA GAACTTGGTG  

AACCACGCCA GTGAGCTTTC GAGAACGGGG ATTATGCTT CTGAAACCC

```

hipDI

```

-----  

L Y A K S E F K D R V V I A K V -  

1151 CTCTGTATGC AAAAACGGAG TTCAAAGACCC GTGTTGTAAT TGCTAAAGTT  

GAGACATACG TTTTTCGCTC AAAGTTCTGG CACAACTTAA ACAGTTTCAA

```

hipDI

```

-----  

D A T A N D V P D E I Q G F P T I -  

1201 GATGCAACAG CTAACGATGT TCCAGATGAA ATTCAAGGGAT TCCCTACTAT  

CTACGGTGTG GATTGCTACA AGGTTCTACTT TAAGTTCCTA AGGGATGATA

```

hipDI

```

-----  

K L Y P A G A K G Q P V T Y S G S -  

1251 CAAACTATAC CCAGCTGGT CAAAGGTCAC ACCTGTACT TACTCTGGTT  

GTTTGATATG GGTGACCCAC GTTTTCCAGT TGACAAATGA ATGAGACCAA

```

hipDI

```

-----  

R T V E D L I K F I A E N G K Y -  

1301 CACCACTATGT TGAAGGCTT ATCAAATTCA TTGCTGAAAA CGGTAAATAC  

GTGCGTGUACA ACTTCTGGAA TAGITTAAGT AACGACTTTTG GCCATTATAG

```

hipDI

```

-----  

S p I
-----  

K A A I S E D A E E T S S A T E T -  

1351 AAAGCTGCAA TCTCAGAAGA TGCTGAAAGG ACTAGTTCAG CAACTGAAAC  

TTTGACGTT AGAGTCTCT ACAGCTCTC TGATCAAGTC GTTGACTTGTG

```

hipDI

```

-----  

T T E T A T K S E A A K E T A T -  

1401 AACTACAGAA ACTGCTACAA AGTCAGAGAA AGCTGCAAAA GAAACTGCAA  

TTGAGTGTCT TGACGATGTT TGAGTCTCTC TGAGCTGTTT CTTTGACGTT

```

Enteropeptidase cleavage linker

hipDI

N-term BBI

```

-----  

E H D E L G S G S G D D D D K K D -  

1451 CAGAACACGA CGAACCTGG TCTGGTCCG GAGATGACCA TGACAAAGAC  

GCTCTGTGCT GCTGAAACT AGACCAAGG CTCTACTGCT ACTGTTCTG

```

N-term BBI

SacI

```

-----  

D E S S -  

1501 GATGAGAGCT CT (SEQ ID NO:5)  

CTACTCTCGA GA (SEQ ID NO:6)

```

FIGURE 8.

aprE promoter

EcoRI

1 GAAATTCTCCA TTTTCTTCCTG CTATCAAAAT AACAGACTCG TGATTTTCCA
CTTAAAGGT AAAAGAAGAC GATAGTTTTA TTGTCTGAGC ACTAAAAGGT

aprE promoter

51 AACGAGCTTT CAAAAAAAGCC TCTGCCCTT GCAAATCGGA TGCCCTGCTA
TTGCTGAAAGA GTTTCCTCGG AGACGGGGAA CGTTTACGCTT ACGGACAGAT

aprE promoter

NotI

101 TAAAAATCCC GATATTGGTT AAACAGCGGC GCAATGGGGG CGGCATCTGA
ATTTTAAAGG CTATAACCAA TTGTGCCCC CGTTACCGCC CGCGTAGAGT

aprE promoter

151 TGCTCTTGTGCT TGGCGAAATGT TCATCTTATT TCTTCTCTCC TCTCAATAAT
ACAGAAACGA ACCGCTTACA AGTAGAATAA AGAAGGGAGGG AGAGTTTATTA

aprE promoter

201 TTTTTCATTC TATCCCTTTT CTGTAAGTT TATTTTTCAAG AATACTTTTA
AAAAAAGTAAG ATAGGGAAAAGA GACATTTCAA AAAAAAAGTC TTATGAAAAT

aprE promoter

251 TCACTATGCT TTGAAAATAAT ATCACGATATA TATCCATTTGT TCTCACGGAA
AGTAGTACGA AACATTTTTA TAGTGCTTATT ATAGGTAACA AGAGTGCCCT

aprE promoter

301 GCACACCCAG GTCATTTGAA CGAAATTTCG CGACAGGAAT TTGCGGGAC
CTGTGCGTC CAGTAAACCA ATTTTGATTGTT GCTGTCCTTA AACGGCCCTG

aprE promoter

351 TCAAGGAGCAT TTAACCTTAA AAAACATGAC ATTTCAAGCAT AATGAAACATT
ATGCTCTGTA AATTGATTGTT TTTCTGACTG TAAAGTCGTA TTACTTGATT

aprE promoter

401 TACTCTGTC TATTTTCGTT CTTTCTGTA TGAAAATAGT TATTTCGAGT
ATGAGTACAG ATAAAAACCAAA GAAAAGACAT ACTTTTATCA ATAAAGCTCA

aprE promoter

451 CTCTAACGGAA ATAGGGAGAG ATGATATACC TAAATAGAGA TAAAATCATC
GAGATGCGCTT TATGCCCTCTC TACTATATGG ATTTATCTCT ATTTTGTAG

aprE promoter

501 TCAAAAAAAAT GGGCTACTA AAATATTAACTT CCATCTTAATA CAATAAATTC
AGTTTTTTAA CCCAGATGAT TTATATAATAA GGTAGATATAAT GTTATTTAAG

aprE promoter

551 ACAGAAATAGT CTTTTAAAGTA AGTCTACTCT GAATTTTTTTT AAAAGGAGAG
TGTCTTATCA GAAAATCTATCA TCAGATGAGA CTTAAAAAAA TTTCCTCTC

aprE promoter Cutinase signal peptide

AprE signal peptide

FIGURE 8 (cont.)

 V R S K K L W I S L L F A L
 601 GTTAAAGAGT GAGAAGCAAA AATTTGTGGA TCAGCTGTTT GTTTGGGTTA
 CCATTTCTCA CTCTTCGTTT TTTAACACCT AGTCGAACAA CAAACCCAT
 Cutinase

 Cutinase signal peptide

 T L A A S C L S V C A T V A A A P .
 651 ACGCTGGCGG CCTCTTGCCT GTCCGCTCTGTG GCACTGTGCG CGCGGGCTCC
 TGCGACCGCC GGAGAACCGGA CAGGACAGACA CGGTGACAGNC GCGCCGGAGG
 Cutinase

 L P D T P F G A P F P A V A N F D R .
 701 CTGCGCGGT ACACCCGACCT GCACTTTTCC CGCTGTGCGCC AATTTGACCC
 TGACGGCGCTA TGTTGGCCCTC GCGGTAAGG CGCACAGCGG TTAAAGCTGG
 Cutinase

 S G P Y T T S S Q S E G P S C R .
 751 GCGATGGCGGC CTACACCAAC AGCACCGAGA CGAGGGGGCC GAGCTGTGCG
 CGTCACCCGGG GATGTGTTGG CGCTCGGTCT CGCTCCCGGG CTCGACAGCG
 Cutinase

 I Y R P R D L G Q G G V R H P V I .
 801 ATCTATCGGC CCCCCGACCT GGCTCAGGGG GCGCTGCGTC ATCCGGTGTAT
 TAGATACCGG GGGCGCTGGA CCGACTCCCC CGCACCGAG TAGGCCACTA
 Cutinase

 L W G N G T G A G P S T Y A G L L .
 851 TCTCTGGGGGC ATATGCCACCG GTGCGGGGSC GTCCACCTAT GCGCGCTTGC
 AGAGACCCCG TTACCGTGCG CACGGCCCGG CAGGTGGATA CGGCCGAACG
 Cutinase

 S H W A S H G F V V A A A E T S .
 901 TATCCGACTG GCGAACCCAC GGTTCTGGGG TGCGGGCGGC GAAACCTCC
 ATAGCGTGAC CGCTTGGTG CCAAGACCC ACCGGCCGCG CCTTTGGAG
 Cutinase

 N A G T G R E M L A C L D Y L V R .
 951 ATAGCGGCGA CGGGCGGGGA AATGCTCCCG TGCTCGGACT ATCTGGTAGG
 TTACCGCCAT GGGCCGCCCT TTACGAGCGG ACGGACCTGA TAGACCATGC
 Cutinase

 E N D T P Y G T Y S G K L N T G R .
 1001 TGAGGACGAC ACCCTCTACG GCACTTATTC CGCAAGCTC AATACCGGC
 ATCTCTGCTG TGGGGATGCG CGTGGATAAG GCGCTTCGAG TTAATGGCCCG
 Cutinase

 V G T S G H S Q G G G G S I M A .
 1051 GAGTCGGCAC TTCTGGGCAT TCCCAAGGGTGTG GTGGCGGCTC GATCATGGCC
 CTCAAGCCGCTG AAGACCCCTA AGGGTCCACG CACGGCCGAG CTAGTACCGC
 Cutinase

 G Q D T R V R T T A P I Q P Y T L .
 1101 GGCAGGATA CGAGGGTGC TACACGGCGG CGGATCCAGC CCTACACCC
 CGCGTCTAT CGTCCACCGC ATGGTGCAGC GGCTAGGTGCG GGGATGTGGGA
 Cutinase

FIGURE 8 (cont.)

1151 . G L G H D S A S Q R R Q Q G P M F .
 CGGCCTGGGG CACGACAGCG CCTGGCAGCG GCGGCAGCG GGGCCGATGT
 GCGGACCCC GTGCTGTCGC GGAGCGTCGC CGCCGTCGTC CCCGGCTACA

Cutinase

1201 . L M S G G G D T I A F P Y L N A .
 TCCCTGATGTC CGGTGGGGT GACACATCG CCTTTCCTCA CCTCAACGCT
 AGGACTACAG GCCACCCCA CTGTGGTAGC GGAAAGGGAT GGAGTTGCGA

Cutinase

1251 . Q P V Y R R A N V P V F W G E R R .
 CAGCCGGTCT ACCCGCGTGC CAAATGTGCCG GTGTTCTGGG GCGAACGGCG
 GTCGGCCAGA TGGCCGACCG GTTACACGGC CACAAGACCC CGCTTGCCGC

Cutinase

1301 . Y V S H F E P V G S G G A Y R G P .
 TTACGTCAGC CACTTCGAGC CGGTGGTAG CGGTGGGGCC TATGGCGGCC
 AATGCAGTCG GTGAAGCTCG GCCRGCCATC GCCACCCCG ATAGCGCCGG

Cutinase

1351 . S T A W F R F Q L M D D Q D A R .
 CGAGCACGGC ATGGTTCCGC TTCCAGCTGA TGATGACCA AGACGCCGC
 GCTCGTGGCG TACCAAGGGC AAGGTGCACT ACCTACTGGT TCTGGGGCG

Cutinase

Alw44I

1401 . A T F Y G A Q C S L C T S L L W S .
 GCTACCTCT ACGGCCGGCA GTGCAGCTCG TGCACTTC TGCTTGGTC
 CGATGGAAGA TGCGCGCGT CACGTAGAC ACGTGAAGAG AGGAACCAG

Linker 2

Cutinase

BamHI

1451 . V E R R G L D N N D P I P D (SEQ ID NO:7)
 TGTTGAAACG AGAGGCTTG ACACAAATGA TCCATTCCG GATCC (SEQ ID NO:8)
 ACAACTTGGC TCTCCAGAAC TTGTGTTACT AGGATAAGGC CTAGG

**BACTERIAL EXPRESSION OF
BOWMAN-BIRK PROTEASE INHIBITOR
AND VARIANTS THEREOF**

The present application claims priority under 35 U.S.C. § 119, to co-pending U.S. Provisional Patent Application Ser. No. 60/518,154, filed Nov. 6, 2003, co-pending U.S. Provisional Patent Application Ser. No. 60/520,403, filed Nov. 13, 2003, co-pending U.S. Provisional Patent Application Ser. No. 60/530,954, filed Dec. 19, 2003, co-pending U.S. Provisional Patent Application Ser. No. 60/531,207, filed Dec. 19, 2003, and co-pending U.S. Provisional Patent Application Ser. No. 60/531,189, filed Dec. 19, 2003.

FIELD OF THE INVENTION

The present invention provides compositions and methods related to expression of protease inhibitors and variants thereof in bacterial species. The present invention further provides fusion nucleic acids, vectors, fusion polypeptides, and processes for obtaining the protease inhibitors.

BACKGROUND OF THE INVENTION

Proteases are involved in a wide variety of biological processes. Disruption of the balance between proteases and protease inhibitors is often associated with pathologic tissue destruction. Indeed, various studies have focused on the role of proteases in tissue injury, and it is thought that the balance between proteases and protease inhibitors is a major determinant in maintaining tissue integrity. Serine proteases from inflammatory cells, including neutrophils, are implicated in various inflammatory disorders, such as pulmonary emphysema, arthritis, atopic dermatitis and psoriasis.

Proteases also appear to function in the spread of certain cancers. Normal cells exist in contact with a complex protein network, called the extracellular matrix (ECM). The ECM is a barrier to cell movement and cancer cells must devise ways to break their attachments, degrade, and move through the ECM in order to metastasize. Proteases are enzymes that degrade other proteins and have long been thought to aid in freeing the tumor cells from their original location by chewing up the ECM. Recent studies have suggested that they may promote cell shape changes and motility through the activation of a protein in the tumor cell membrane called Protease-Activated Receptor-2 (PAR2). This leads to a cascade of intracellular reactions that activates the motility apparatus of the cell. Thus, it is hypothesized that one of the first steps in tumor metastasis is a reorganization of the cell shape, such that it forms a distinct protrusion at one edge facing the direction of migration. The cell then migrates through a blood vessel wall and travels to distal locations, eventually reattaching and forming a metastatic tumor. For example, human prostatic epithelial cells constitutively secrete prostate-specific antigen (PSA), a kallikrein-like serine protease, which is a normal component of the seminal plasma. The protease acts to degrade the extracellular matrix and facilitate invasion of cancerous cells.

Synthetic and natural protease inhibitors have been shown to inhibit tumor promotion in vivo and in vitro. Previous investigations have indicated that certain protease inhibitors belonging to a family of structurally-related proteins classified as serine protease inhibitors or SERPINS, are known to inhibit several proteases including trypsin, cathepsin G, thrombin, and tissue kallikrein, as well as neutrophil elastase. The SERPINS are extremely effective at preventing/suppressing carcinogen-induced transformation in vitro and car-

cinogenesis in animal model systems. Systemic delivery of purified protease inhibitors apparently reduces joint inflammation and cartilage and bone destruction as well.

Topical administration of protease inhibitors finds use in such conditions as atopic dermatitis, a common form of inflammation of the skin, which may be localized to a few patches or involve large portions of the body. The depigmenting activity of protease inhibitors and their capability to prevent ultraviolet-induced pigmentation have been demonstrated both in vitro and in vivo (See e.g., Payne et al., *J. Invest. Dermatol.*, 116:587-595 [2001]). Protease inhibitors have also been reported to facilitate wound healing. For example, secretory leukocyte protease inhibitor was demonstrated to reverse the tissue destruction and speed the wound healing process when topically applied. In addition, serine protease inhibitors can also help to reduce pain in lupus erythematosus patients (See e.g., U.S. Pat. No. 6,537,968).

As noted above, protease inhibitors interfere with the action of proteases. Naturally occurring protease inhibitors can be found in a variety of foods such as cereal grains (oats, barley, and maize), Brussels sprouts, onion, beetroot, wheat, finger millet, and peanuts. One source of interest is the soybean. The average level of protease inhibitors present in soybeans is around 1.4 percent and 0.6 percent for Kunitz and Bowman-Birk, respectively, two of the most important protease inhibitors. Notably, these low levels make it impractical to isolate the natural protease inhibitor for clinical applications.

Thus, there is a need for methods and compositions suitable for the large-scale production of protease inhibitors and their variants. In particular, there remains a need for compositions and methods that reduce and/or eliminate risks associated with blood-borne infectious agents when these proteases are produced in mammalian tissue culture cells.

SUMMARY OF THE INVENTION

The present invention provides compositions and methods related to expression of protease inhibitors and variants thereof in bacterial species. The present invention further provides fusion nucleic acids, vectors, fusion polypeptides, and processes for obtaining the protease inhibitors.

Provided herein are nucleic acids, cells and methods for the production of protease inhibitors and variants thereof.

The present invention provides nucleic acids encoding at least one functional protease inhibitor. In one aspect, a nucleic acid comprising regulatory sequences operatively linked to a first, second, third and fourth nucleic acid sequences are provided. In some embodiments, terminator sequences are provided following the fourth nucleic acid sequence.

In alternative embodiments, the first nucleic acid sequence encodes a signal polypeptide functional as a secretory sequence in a first host organism, the second nucleic acid encodes a secreted polypeptide or functional portion thereof normally secreted from the first or a second host organism, the third nucleic acid encodes a cleavable linker and the fourth nucleic acid encodes a protease inhibitor or fragment thereof. In further embodiments, the present invention provides at least one expression cassette comprising nucleic acid sequences encoding at least one protease inhibitor.

In additional embodiments, the present invention provides polynucleotides encoding at least one protease inhibitor variant. In some particularly preferred embodiments, the polynucleotide encodes a Bowman-Birk Inhibitor (BBI) variant, wherein at least one loop of the wild-type BBI has been altered.

The present invention also provides methods of expressing functional protease inhibitors or variants thereof. In some preferred embodiments, host cells suitable for production of functional protease inhibitors and/or variants thereof are provided. In some embodiments, a host cell is (i) transformed with at least one expression cassette comprising a nucleic acid sequence encoding at least one protease inhibitor or variant thereof, and (ii) cultured under appropriate conditions to express at least one protease inhibitor or variants thereof. In some embodiments of the methods, the method further comprises recovering the protease inhibitor or variant thereof.

In alternative embodiments, a host cell is (i) transformed with a first expression cassette comprising at least one nucleic acid sequence encoding a protease inhibitor or variant thereof, (ii) transformed with a second expression cassette comprising a nucleic acid sequence encoding at least one thiol-disulfide oxidoreductase or chaperone, and (iii) cultured under appropriate conditions to express the protease inhibitors or variant thereof. In some embodiments, the protease inhibitors or variants thereof are recovered. In some preferred embodiments, the protease inhibitors or variant thereof are expressed as a fusion protein. In further embodiments, the methods further comprise recovering the protease inhibitor or variant thereof.

In still further embodiments, functional protease inhibitors and variants thereof are provided. In some embodiments, the functional protease inhibitor or variant thereof is expressed as a fusion protein consisting of the signal sequence, a cellulase catalytic domain, a cleavable linker region, and then by the mature protease inhibitor or variant thereof.

In additional embodiments, the expressed proteins are treated with a protease and/or acid/heat treatment to liberate a protease inhibitor or variant thereof from the fusion protein.

In further embodiments, the present invention provides a polypeptide having protease inhibitory activity, selected from the group consisting of

a) Bowman-Birk Inhibitor variants;

b) Bowman-Birk Inhibitor; and

c) a scaffold comprising at least one variant sequence.

The present invention provides compositions comprising a fusion protein, wherein the fusion protein comprises a protease inhibitor, and a peptide of interest. In some embodiments, the fusion protein comprises an amino acid selected from the group consisting of SEQ ID NOS:2, 4, 6 and 8. In alternative embodiments, the fusion protein is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, and 7. In some preferred embodiments, the protease inhibitor is selected from the group consisting of Bowman-Birk inhibitor (BBI), soybean trypsin inhibitor (STI), and egrin C. In some particularly preferred embodiments, the protease inhibitor is BBI, and wherein the BBI comprises at least one loop selected from the group consisting of a trypsin loop and a chymotrypsin loop. In further preferred embodiments, the protease inhibitor is a scaffold for the peptide of interest. In some preferred embodiments, the loop comprises at least one peptide of interest. In some particularly preferred embodiments, the peptide of interest comprises the amino acid sequence set forth in SEQ ID NO:9. In some more preferred embodiments, the fusion protein comprises the amino acid sequence set forth in SEQ ID NO:4.

The present invention also provides methods for producing at least one protease inhibitor in a bacterial cell comprising: a) introducing a DNA construct into a bacterial cell, wherein the DNA construct comprises a heterologous DNA sequence encoding a protease inhibitor derived from a Bowman-Birk inhibitor (BBI) or variants thereof; b) culturing the bacterial cell under suitable culture conditions to allow expression of

the heterologous DNA sequence; and c) producing the protease inhibitor. In some embodiments, the construct further comprises at least one catalytic domain. In some embodiments, catalytic domain is selected from the group consisting of cellulase, cutinase, and disulfide isomerase. In some preferred embodiments, the method further comprises the step of recovering the protease inhibitor. In some particularly preferred embodiments, the method further comprises the step of activating the protease inhibitor. In some most preferred

embodiments, the activating is accomplished by exposing the protease inhibitor to at least one reagent selected from the group consisting of compositions that reduce disulfide bonds, compositions that oxidize disulfide bonds, and compositions that alter the redox potential. In some embodiments, the bacterial cell is a member of the genus *Bacillus*. In additional

embodiments, protease inhibitor is selected from the group consisting of: i) a protease inhibitor having at least 90% sequence identity with SEQ ID NO: 11; and ii) a protease inhibitor having at least 90% sequence identity with SEQ ID NO:13. In yet further embodiments, the protease inhibitor comprises a variant sequence. In additional embodiments, the protease inhibitor is Bowman-Birk inhibitor and wherein SEQ ID NO:9 is substituted for at least one loop, wherein the loop is selected from the group consisting of the trypsin loop

and the chymotrypsin loop. In some embodiments, the method further comprises the step of introducing a second nucleic acid sequence encoding a thiol-disulfide oxidoreductase or chaperone into the bacterial cell. In some preferred

embodiments, the protease inhibitor is expressed as a fusion protein. In some preferred embodiments, the fusion protein further comprises a cellulase catalytic domain, a cleavage site, and the protease inhibitor. In some particularly preferred

embodiments, the fusion protein is processed by a protease or acid/heat treatment to liberate the protease inhibitor. In still

further preferred embodiments, the fusion protein further comprises at least one linker sequence. In some embodiments, the linker sequence is selected from the group consisting of SEQ ID NOS:141-143.

The present invention also provides protease inhibitor

compositions comprising the protease inhibitor produced according to any and all of the methods described herein. In some embodiments, the various primers and oligonucleotides described herein find use in the production of the fusion proteins of the present invention. In addition, the present invention provides numerous peptides suitable for use in the present invention.

The present invention also provides methods for inhibiting the proteolytic activity of a target protein comprising contacting the target protein with any of the protease inhibitor compositions described herein, and binding the target protein wherein the proteolytic activity of the target protein is inhibited.

The present invention further provides isolated polynucleotides encoding a protease inhibitor selected from the group

consisting of polypeptide sequences set forth in SEQ ID NOS:10 and 12.

The present invention also provides expression vectors comprising a polynucleotide sequence, wherein the polynucleotide sequence is selected from the following: a) a polynucleotide sequence encoding a protease inhibitor having at least 90% sequence identity with SEQ ID NO: 11, wherein at least one of the loop residues have been replaced with a variant sequence; and b) a polynucleotide sequence encoding a protease inhibitor having at least 90% sequence identity with SEQ ID NO: 13, wherein at least one of the loop residues have been replaced with a variant sequence. In some embodiments, the expression vectors further comprise the 5'

terminus to the 3' terminus: a first nucleic acid sequence encoding a signal peptide functional as a secretory sequence in a bacterial cell; a second nucleic acid sequence encoding a secreted polypeptide or functional portion thereof; a third nucleic acid sequence encoding a cleavable linker; and the DNA sequence which encodes the protease inhibitor. The present invention also provides host cells transformed with the expression vectors of the present invention. In some preferred embodiments, the host cell is a *Bacillus* species cell.

The present invention also provides compositions comprising at least one scaffold protein and at least one peptide, wherein the scaffold comprises Bowman-Birk inhibitor. In some preferred embodiments, the peptide comprises a peptide that binds to VegF.

Other objects, features and advantages of the present invention are apparent from the following detailed description. It should be understood, however, that the detailed description and specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the scope and spirit of the invention will be apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 provides the DNA and amino acid sequences of the aprl-BCE103-BB1-Histag expression cassette (EcoRI-Hin-dII) cloned into the pJM103 integration vector (SEQ ID NOS:1 and 2).

FIG. 2 provides a schematic map of the pJM103BB1his expression vector.

FIG. 3 provides the DNA and amino acid sequences of 2BBlck81 from the BCE103 fusion site (at the BamHI) to the end of the gene (SEQ ID NOS:3 and 4). The CK37281 peptide sequences (ACYNLYGW1C' (SEQ ID NO:9) are inserted into both the trypsin and chymotrypsin inhibitory loops.

FIG. 4 provides a graph showing titers of active versus inactive 2BBlck81 (by trypsin inhibition) and the ratio of the activities of BCE103 cellulase to 2BBlck81 with various thiol reducing agents added during the growth of the culture. In this Figure, BME=2-mercaptoethanol, Cyt=cysteine, Glut=reduced glutathione, DTT=dithiothreitol.

FIG. 5 provides a graph showing activation of BCE-Ink2-2BBlck81 with 2-mercaptoethanol (bME) after partial purification by ion exchange chromatography.

FIG. 6 provides a graph showing results from a competition analysis of 2BBlck81 versus anti-VegF antibody binding to VegF.

FIG. 7 provides the sequence of the synthetic DNA fragment carrying the *H. insolens* PDI (hilPDI) that was inserted into the *B. subtilis* BHI expression vector, as well as the amino acid sequence (SEQ ID NOS:5 and 6).

FIG. 8 provides the DNA and amino acid sequences of the aprl-cutinase expression cassette that was ligated into the EcoRI-BamHI sites of p2JM103-Ink2-2BBlck81 (SEQ ID NOS:7 and 8).

DESCRIPTION OF THE INVENTION

The present invention provides compositions and methods related to expression of protease inhibitors and variants thereof in bacterial species. The present invention further provides fusion nucleic acids, vectors, fusion polypeptides, and processes for obtaining the protease inhibitors.

Unless otherwise indicated, the practice of the present invention involves conventional techniques commonly used

in molecular biology, microbiology, and recombinant DNA, which are within the skill of the art. Such techniques are known to those of skill in the art and are described in numerous texts and reference works (See e.g., Sambrook et al., "Molecular Cloning: A Laboratory Manual", Second Edition (Cold Spring Harbor), [1989]; and Ausubel et al., "Current Protocols in Molecular Biology" [1987]). All patents, patent applications, articles and publications mentioned herein, both supra and infra, are hereby expressly incorporated herein by reference.

Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. For example, Singleton and Sainsbury,

15 *Dictionary of Microbiology and Molecular Biology*, 2d Ed, John Wiley and Sons, NY (1994); and Hale and Marham, *The Harper Collins Dictionary of Biology*, Harper Perennial, N.Y. (1991) provide those of skill in the art with a general dictionaries of many of the terms used in the invention. Although

20 any methods and materials similar or equivalent to those described herein find use in the practice of the present invention, the preferred methods and materials are described herein. Accordingly, the terms defined immediately below are more fully described by reference to the Specification as a whole. Also, as used herein, the singular "a", "an" and "the" includes the plural reference unless the context clearly indicates otherwise. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid

25 sequences are written left to right in amino to carboxy orientation, respectively. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art.

30 Furthermore, the headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

35 Nonetheless, in order to facilitate understanding of the invention, a number of terms are defined below.

Definitions

As used herein, the terms "expression cassette" and "expression vector" refer to nucleic acid constructs generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a target cell. The recombinant expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of an expression vector includes, among other sequences, a nucleic acid sequence to be transcribed and a promoter. The term "expression cassette" may be used interchangeably herein with "DNA construct" and its grammatical equivalents.

45 As used herein, the terms "vector" and "cloning vector" refer to nucleic acid constructs designed to transfer nucleic acid sequences into cells.

50 As used herein, the term "expression vector" refers to a vector that has the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those of skill in the art.

55 As used herein, the term "plasmid" refers to a circular double-stranded (ds) DNA construct used as a cloning vector,

60

and which forms an extrachromosomal self-replicating genetic element in some eukaryotes or prokaryotes, or integrates into the host chromosome.

As used herein, the terms "nucleic acid molecule" and "nucleic acid sequence" include sequences of any form of nucleic acid, including, but not limited to RNA, DNA and cDNA molecules. It will be understood that, as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding a given protein may be produced.

As used herein, a "fusion DNA sequence" comprises from 5' to 3' a first, second, third and fourth DNA sequences.

As used herein, "first nucleic acid sequence" or "first DNA sequence" encodes a signal peptide functional as a secretory sequence in bacterial species. Particularly preferred signal sequences are those derived from polypeptides secreted by the expression host used to express and secrete the fusion polypeptide. As used herein, first amino acid sequences correspond to secretory sequences which are functional in a bacterial species. Such amino acid sequences are encoded by first DNA sequences as defined.

As used herein, "second DNA sequences" encode "secreted polypeptides" (i.e., "secreted polypeptides of interest") expressed by the chosen bacterial species. As with the first DNA sequences, preferred secreted polypeptides are those which are naturally secreted by the bacterial expression host. However, in some embodiments, the term refers to heterologous protein (i.e., proteins that are not normally secreted by the particular bacterial host). For example, in some preferred embodiments, a cellulase expressed by a *Bacillus* species other than *B. subtilis* is used that the secreted polypeptide of interest and *B. subtilis* is used as the expression host.

As used herein, "functional portion of a secreted polypeptide" and its grammatical equivalents refers to a truncated secreted polypeptide that retains its ability to fold into a normal, albeit truncated, configuration. In some embodiments, it is contemplated that sufficient residues of a domain of the secreted polypeptide must be present to allow it to fold in its normal configuration independently of the desired polypeptide to which it is attached. However, in most cases, the portion of the secreted polypeptide are both correctly folded and result in increased secretion as compared to its absence.

Similarly, in most cases, the truncation of the secreted polypeptide means that the functional portion retains a biological function. In a preferred embodiment, the catalytic domain of a secreted polypeptide is used, although other functional domains may be used, for example, the substrate binding domains. Additionally preferred embodiments utilize the catalytic domain and all or part of the linker region.

As used herein, "third DNA sequences" comprise DNA sequences encoding a cleavable linker polypeptide. It should be understood that the third DNA sequence need only encode that amino acid sequence which is necessary to be recognized by a particular enzyme or chemical agent to bring about cleavage of the fusion polypeptide. Thus, only that portion of the linker which is necessary for recognition and cleavage by the appropriate enzyme is required.

As used herein, "fourth DNA sequences" encode "desired polypeptides." Such desired polypeptides include protease inhibitors either in their mature or pro forms, and variants thereof.

The above-defined four DNA sequences encoding the corresponding four amino acid sequences are combined to form a "fusion DNA sequence." Such fusion DNA sequences are assembled in proper reading frame from the 5' terminus to 3' terminus in the order of first, second, third and fourth DNA sequences. As so assembled, the DNA sequence encodes a

"fusion polypeptide," "fusion protein," and "fusion analog" encoding from its amino-terminus a signal peptide functional as a secretory sequence in a bacterial species, a secreted polypeptide or portion thereof normally secreted by a bacterial species, a cleavable linker polypeptide and a desired polypeptide.

As used herein, the terms "desired protein" and "desired polypeptide" refer to a polypeptide or protein in its mature or pro form that is not fused to a secretion enhancing construct. Thus, a "desired protein" and "desired polypeptide" refer to the protein to be expressed and secreted by the host cell in a non-fused form.

As used herein, "fusion polypeptides," "fusion proteins," and "fusion analogs" encode from the amino-terminus a signal peptide functional as a secretory sequence functional in a host cell, a secreted polypeptide or portion thereof normally secreted from a host cell, a cleavable linker polypeptide and a desired polypeptide. In some embodiments, the fusion protein is processed by host cell enzymes (e.g., a protease), to yield the desired protein free from the other protein sequences in the fusion protein. As used herein, the terms "fusion analog," "fusion polypeptide," and "fusion protein" are used interchangeably.

As used herein, a "promoter sequence" refers to a DNA sequence which is recognized by the bacterial host for expression purposes. In preferred embodiments, it is operably linked to a DNA sequence encoding the fusion polypeptide. Such linkage comprises positioning of the promoter with respect to the translation initiation codon of the DNA sequence encoding the fusion DNA sequence. In particularly preferred embodiments, the promoter sequence contains transcription and translation control sequences which mediate the expression of the fusion DNA sequence.

As used herein, "terminator sequence" refers to a DNA sequence which is recognized by the expression host to terminate transcription. It is operably linked to the 3' end of the fusion DNA encoding the fusion polypeptide to be expressed.

As used herein, the term "selectable marker-encoding nucleotide sequence" refers to a nucleotide sequence which is capable of expression in bacterial cells and where expression of the selectable marker confers to cells containing the expressed gene the ability to grow in the presence of a corresponding selective condition.

As used herein, a nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA encoding a secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Operably linked DNA sequences are usually contiguous, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

As used herein, "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid sequence or that the cell is derived from a heterologous cell or modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native

genes that are otherwise abnormally expressed, under expressed or not expressed at all as a result of deliberate human intervention.

As used herein, the term "expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation. Thus, the term "protease inhibitor expression" refers to transcription and translation of the specific protease inhibitors and variants thereof gene to be expressed, the products of which include precursor RNA, mRNA, polypeptide, post-translation processed polypeptide, and derivatives thereof. Similarly, "protease inhibitor expression" refers to the transcription, translation and assembly of protease inhibitors and variants thereof exemplified by the expression cassette shown in FIG. 1. By way of example, assays for protease inhibitor expression include examination of bacterial colonies when exposed to the appropriate conditions, Western blots for protease inhibitor protein, as well as Northern blots analysis and reverse transcriptase polymerase chain reaction (RT-PCR) assays for protease inhibitor mRNA.

As used herein, the terms "isolated" and "purified" as used herein refer to a nucleic acid or polypeptide that is removed from at least one component with which it is naturally associated.

As used herein, the term "substantially free" encompasses preparations of the desired polypeptide having less than about 20% (by dry weight) other proteins (i.e., contaminating protein), less than about 10% other proteins, less than about 5% other proteins, or less than about 1% other proteins.

As used herein, the term "substantially pure" when applied to the proteins or fragments thereof of the present invention means that the proteins are essentially free of other substances to an extent practical and appropriate for their intended use. In particular, the proteins are sufficiently pure and are sufficiently free from other biological constituents of the host cells so as to be useful in, for example, protein sequencing, and/or producing pharmaceutical preparations.

As used herein, the term "target protein" refers to protein (e.g., enzyme, hormone, etc.), whose action would be blocked by the binding of the variant inhibitors provided for herein.

As used herein, the terms "variant sequence" and "variant sequences" refer to the short polypeptide sequence(s) that replace the binding loops of the wild-type protease inhibitor or other scaffold. The variant sequence does not need to be of the same length as the binding loop sequence it is replacing on the scaffold.

As used herein, the term "scaffold" refers to a wild-type protein sequence into which a variant sequence is introduced. In some embodiments, the scaffold has portions (e.g., loops), that are replaced. For example, the BBI sequences provided herein find use as scaffolds for variant sequences.

Protease Inhibitors

Two protein protease inhibitors have been isolated from soybeans, the Kunitz-type trypsin inhibitor (soybean trypsin inhibitor, STI) and the Bowman-Birk protease inhibitor (BBI) (See e.g., Birk, Int. J. Pept. Protein Res., 25:113-131 [1985]; and Kennedy, Am. J. Clin. Nutr., 68:1406S-1412S [1998]). These inhibitors serve as a scaffold for the variant sequences. In addition to alterations in the scaffold comprising the variant sequences, other desired proteins used herein include the addition of six histidine residues at the C-terminus (See, FIGS. 1 and 2).

Soybean Trypsin Inhibitor (STI)

STI inhibits the proteolytic activity of trypsin by the formation of a stable stoichiometric complex (See e.g., Liu,

Chemistry and Nutritional Value of Soybean Components. In: *Soybeans, Chemistry Technology and Utilization*, pp. 32-35, Aspen Publishers, Inc., Gaithersburg, Md., [1999]). STI consists of 181 amino acid residues with two disulfide bridges

and is roughly spherically shaped (See e.g., Song et al., J. Mol. Biol., 275:347-63 [1998]). The trypsin inhibitory loop lies within the first disulfide bridge. The Kunitz-type soybean trypsin inhibitor (STI) has played a key role in the early study of proteinases, having been used as the main substrate in the biochemical and kinetic work that led to the definition of the standard mechanism of action of proteinase inhibitors.

Bowman-Birk Inhibitor (BBI)

Bowman-Birk inhibitor proteins are a kinetically and structurally well-characterized family of small proteins (60-90 residues) isolated from leguminous seeds, as well as other plants, including various grasses. They typically have a symmetrical structure of two tricyclic domains each containing an independent binding loop, although some have one domain and some have more than two domains. The major ~8 kDa Bowman-Birk inhibitor isolated from soybeans (BBI) has two separate reactive site loops, loop I inhibits proteinases having trypsin-like specificity and loop II inhibits proteinases with chymotrypsin-like specificity (See e.g., Chen et al., J. Biol. Chem., 267:1990-1994 [1992]; Werner and Wemmer, Biochem., 31:999-1010 [1992]; Liu et al., Eur. J. Biochem., 212:549-555 [1993]; Voss et al., Eur. J. Biochem., 242:122-131 [1996]; and Billings et al., Proc. Natl. Acad. Sci., 89:3120-3124 [1992]). These binding regions each contain a "cannonical loop" structure, which is a motif found in a variety of serine proteinase inhibitors (Bode and Huber, Eur. J. Biochem., 204:433-451 [1992]). STI and BBI are found only in the soybean seed, and not in any other part of the plant (See e.g., Birk, Int. J. Pept. Protein Res., 25:113-131 [1985]).

Although numerous isoforms of BBI have been characterized, SEQ ID NO:13 shows the amino acid sequence of the BBI backbone used herein comprising approximately 71 amino acid residues (See Example 1).

In soybeans, BBI is produced as a pro-protein with an N-terminal pro-peptide that is 19 amino acids in length. Thus, in some embodiments, BBI is produced with all or at least a portion of the propeptide. In some embodiments, BBI is truncated, with as many as 10 amino acid residues being removed from either the N- or C-terminal. For example, upon seed desiccation, some BBI molecules have the C-terminal 9 or 10 amino acid residues removed. Thus, proteolysis is generally highly tolerated prior to the initial disulfide and just after the terminal disulfide bond, the consequences of which are usually not detrimental to the binding to target protein. However, it will be appreciated that any one of the isoforms or truncated forms find use in various embodiments of the present invention.

Protease Inhibitor Variants

As indicated above, the STI and BBI protease inhibitors have binding loops that inhibit proteinases. The present invention provides protease inhibitor variants with alterations in one or more reactive sites (e.g., Loop I and/or Loop II of BBI). In some preferred embodiments, the loops are replaced with sequences that interact with a target protein.

For example, in some embodiments, the loops are replaced with sequences derived from VEGF binding proteins, inhibitors of the complement pathway such as C2, C3, C4 or C5 inhibitors, Comstatin, cytokines, other proteins of interest, etc. Indeed, it is not intended that the present invention be limited to any particular sequence substituted into either of these loops, as any suitable sequence finds use in the present invention.

In some embodiments, variant sequences are selected by various methods known in the art, including but not limited to phage display and other suitable screening methods. For example, a random peptide gene library is fused with phage PIII gene so the peptide library will be displayed on the surface of the phage. Subsequently, the phage display library is exposed to the target protein and washed with buffer to remove non-specific binding (this process is sometimes referred to as panning). Finally, the binding phage and PCR the DNA sequence for the peptide encoded are isolated.

In most embodiments, one of the loops is replaced with a variant sequence (i.e., peptides; often 3 to 14 amino acids in length, with 5 to 10 amino acids being preferred). Longer sequences find use in the present invention, as long as they provide the binding and/or inhibition desired. In addition, peptides suitable for use as replacements of the binding loop(s) preferably adopt a functional conformation when contained within a constrained loop (i.e., a loop formed by the presence of a disulfide bond between two cysteine residues). In some specific embodiments, the peptides are between 7 and 9 amino acids in length. These replacement sequences also provide protease inhibition or binding to the targeted proteins. In some embodiments, it is advantages to alter a single amino acid.

Fusion Proteins

In preferred embodiments, each protease inhibitor or variant thereof is expressed as a fusion protein by the host bacterial cell. Although cleavage of the fusion polypeptide to release the desired protein will often be useful, it is not necessary. Protease inhibitors and variants thereof expressed and secreted as fusion proteins surprisingly retain their function.

The above-defined DNA sequences encoding the corresponding amino acid sequences are combined to form a "fusion DNA sequence." Such fusion DNA sequences are assembled in proper reading frame from the 5' terminus to 3' terminus in the order of first, second, third and fourth DNA sequences. As so assembled, the DNA sequence encodes a "fusion polypeptide" encoding from its amino-terminus a signal peptide functional as a secretory sequence in a bacterial species, a secreted polypeptide or portion thereof normally secreted from a bacterial species, a cleavable linker peptide and a desired polypeptide (e.g., a protease inhibitor and variants thereof). Various methods are known to those in the art for the production of fusion proteins (See e.g., U.S. Pat. Nos. 5,411,873, 5,429,950, and 5,679,545, all of which are incorporated by reference herein). Thus, it is intended that any suitable method will find use in the present invention.

Expression of Recombinant Protease Inhibitors

To the extent that the present invention depends on the production of fusion proteins, it relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook et al., *Molecular Cloning. A Laboratory Manual* (2nd ed.) [1989]; Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and Ausubel et al., (eds.), *Current Protocols in Molecular Biology* (1994).

The present invention provides bacterial host cells which have been transduced, transformed or transfected with an expression vector comprising a protease inhibitor-encoding nucleic acid sequence. The culture conditions, such as temperature, pH and the like, are those previously used for the parental host cell prior to transduction, transformation or transfection are apparent to those skilled in the art.

Basically, a nucleotide sequence encoding a fusion protein is operably linked to a promoter sequence functional in the host cell. This promoter-gene unit is then typically cloned into

intermediate vectors before transformation into the host cells for replication and/or expression. These intermediate vectors are typically prokaryotic vectors (e.g., plasmids, or shuttle vectors). However, it is not intended that the present invention be limited to the use of intermediate vectors, as this step is omitted in some preferred embodiments.

In one approach, a bacterial culture is transformed with an expression vector having a promoter or biologically active promoter fragment or one or more (e.g., a series of) enhancers which functions in the host cell, operably linked to a nucleic acid sequence encoding a protease inhibitor, such that the a protease is expressed in the cell. In some preferred embodiments, the DNA sequences encode a protease inhibitor or variant thereof. In another preferred embodiment, the promoter is a regulatable one.

Nucleic Acid Constructs/Expression Vectors

Natural or synthetic polynucleotide fragments encoding a protease inhibitor (i.e., "PI-encoding nucleic acid sequences") may be incorporated into heterologous nucleic acid constructs or vectors, capable of introduction into, and replication in, a bacterial cell. The vectors and methods disclosed herein are suitable for use in various host cells for the expression of protease inhibitors and variants thereof. Any vector may be used as long as it is replicable and viable in the cells into which it is introduced. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. Appropriate cloning and expression vectors are also described in various references known to those in the art (See e.g., Sambrook et al., *supra* and Ausubel et al., *supra*, expressly incorporated by reference herein). The appropriate DNA sequence is inserted into a plasmid or vector (collectively referred to herein as "vectors") by any suitable method. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by standard procedures known to those in the art.

Appropriate vectors are typically equipped with a selectable marker-encoding nucleic acid sequence, insertion sites, and suitable control elements, such as termination sequences. In some embodiments, the vectors comprise regulatory sequences, including, for example, control elements (i.e., promoter and terminator elements or 5' and/or 3' untranslated regions), effective for expression of the coding sequence in host cells (and/or in a vector or host cell environment in which a modified soluble protein coding sequence is not normally expressed), operably linked to the coding sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, many of which are commercially available and known to those in the art.

Exemplary promoters include both constitutive promoters and inducible promoters. Such promoters are well known to those of skill in the art. Those skilled in the art are also aware that a natural promoter can be modified by replacement, substitution, addition or elimination of one or more nucleotides without changing its function. The practice of the present invention encompasses and is not constrained by such alterations to the promoter. The choice of promoter used in the genetic construct is within the knowledge of one skilled in the art.

The choice of the proper selectable marker will depend on the host cell. Appropriate markers for different bacterial hosts are well known in the art. Typical selectable marker genes encode proteins that (a) confer resistance to antibiotics or other toxins (e.g., ampicillin, methotrexate, tetracycline, neomycin, mycophenolic acid, puromycin, zeomycin, or hydrogycin; or (b) complement an auxotrophic mutation or a naturally occurring nutritional deficiency in the host strain.

In some embodiments, a selected PI coding sequence is inserted into a suitable vector according to well-known recombinant techniques and used to transform a cell line capable of PI expression. Due to the inherent degeneracy of the genetic code, other nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used to clone and express a specific protease inhibitor, as further detailed above. Therefore it is appreciated that such substitutions in the coding region fall within the sequence variants covered by the present invention. Any and all of these sequence variants can be utilized in the same way as described herein for a parent PI-encoding nucleic acid sequence. Those skilled in the art recognize that differing PIs will be encoded by differing nucleic acid sequences.

In some embodiments, once the desired form of a protease inhibitor nucleic acid sequence, homologue, variant or fragment thereof, is obtained, it is modified by any number of ways. Where the sequence involves non-coding flanking regions, the flanking regions may be subjected to resection, mutagenesis, etc. Thus, transitions, transversions, deletions, and insertions may be performed on the naturally occurring sequence.

In some preferred embodiments, heterologous nucleic acid constructs include the coding sequence for at least one protease inhibitor, or variant(s), fragment(s) or splice variant(s) thereof: (i) in isolation; (ii) in combination with additional coding sequences, such as fusion protein or signal peptide coding sequences, where the PI coding sequence is the dominant coding sequence; (iii) in combination with non-coding sequences, such as control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host; and/or (iv) in a vector or host environment in which the PI coding sequence is a heterologous gene.

In some embodiments, heterologous nucleic acid containing the appropriate nucleic acid coding sequence, together with appropriate promoter and control sequences, is employed to introduce into bacterial host cells to permit the cells to express at least one protease inhibitor or variant thereof.

In some embodiments of the present invention, a heterologous nucleic acid construct is employed to transfer a PI-encoding nucleic acid sequence into a cell in vitro. In some preferred embodiments, the host cells stably integrate the nucleic acid sequences of the present invention. Thus, any suitable method for effectively generating stable transformants finds use in the present invention.

In additional embodiments of the present invention, the first and second expression cassettes are present on a single vector, while in other embodiments these cassettes are present on separate vectors.

In some preferred embodiments, in addition to a promoter sequence, the expression cassette also contains a transcription termination region downstream of the structural gene to provide for efficient termination. In some embodiments, the termination region is obtained from the same gene as the promoter sequence, while in other embodiments it is obtained from another gene. The selection of suitable transcription termination signals is well-known to those of skill in the art.

In addition, it is contemplated that any suitable expression vector will find use in the present invention. Indeed, it is contemplated that various conventional vectors used for expression in eukaryotic or prokaryotic cells will be suitable and find use with the present invention. Standard bacterial expression vectors include bacteriophages λ and M13, as well as plasmids such as pBR322 based plasmids, pSKF, pET23D, and fusion expression systems such as MBP, GST, and LacZ.

In further embodiments, epitope tags are added to recombinant proteins, in order to provide convenient methods of isolation (e.g., c-myc).

Additional elements typically included in expression vectors are replicons, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of heterologous sequences. The particular antibiotic resistance gene chosen is not critical, any of the many resistance genes known in the art are suitable.

Introduction of a Protease Inhibitor-Encoding Nucleic Acid Sequence into Host Cells.

In some preferred embodiments, the methods of the present invention provide host cells that contain a stably integrated sequence of interest (i.e., PI-encoding nucleic acid). However, in alternative embodiments, the methods of the present invention provide for maintenance of a self-replicating extrachromosomal transformation vector.

The invention further provides cells and cell compositions which have been genetically modified to comprise an exogenously provided PI-encoding nucleic acid sequence. In some embodiments, a parental host cell is genetically modified by an expression vector. In some embodiments, the vector is a plasmid, while in other embodiments the vector is a viral particle, phage, naked DNA, etc. Thus, it is not intended that the form of the vector be limited to any particular type of vector, as various vectors will find use in the present invention.

Various methods may be employed for delivering an expression vector into cells *in vitro*. Methods of introducing nucleic acids into cells for expression of heterologous nucleic acid sequences are also known to the ordinarily skilled artisan, including, but not limited to electroporation; protoplast fusion with intact cells; transduction; high velocity bombardment with DNA-coated microprojectiles; infection with modified viral (e.g., phage) nucleic acids; chemically-mediated transformation, competence, etc. In addition, in some embodiments, heterologous nucleic acid constructs comprising a PI-encoding nucleic acid sequence are transcribed *in vitro*, and the resulting RNA introduced into the host cell by any of the suitable methods known in the art.

Following introduction of a heterologous nucleic acid construct comprising the coding sequence for a protease inhibitor, the genetically modified cells are cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, and/or amplifying expression of a PI-encoding nucleic acid sequence. The culture conditions, such as temperature, pH and the like, are those previously used for the host cell selected for expression, and are apparent to those skilled in the art.

The progeny of cells into which such heterologous nucleic acid constructs have been introduced are generally considered to comprise the PI-encoding nucleic acid sequence found in the heterologous nucleic acid construct.

Bacterial Hosts and Expression

Appropriate host cells include any suitable bacterial species. In some embodiments, the bacterial hosts serve both as the expression hosts and the source of the first and second nucleic acids. Using the present inventive methods and host cells, surprising levels of expression have been obtained. The system utilized herein has achieved levels of expression and secretion of greater than 0.5 g/l of protease inhibitor.

After the expression vector is introduced into the host cells, the transfected host cells are cultured under conditions favoring expression of gene encoding the desired protein. Large

batches of transformed cells can be cultured as described above. Finally, product is recovered from the culture using techniques known in the art.

Accessory proteins such as thiol-disulfide oxidoreductases or chaperones find use in some embodiments, as they may be beneficial to help fold the secretory protein into its active conformation. Thiol-disulfide oxidoreductases and protein disulfide isomerases catalyze the formation of the correct disulfide bonds in the protein. Overexpression of the bhdC operon in *B. subtilis* has been shown to be beneficial for the production of a protein with disulfide bonds (See e.g., Meima et al., *J. Biol. Chem.*, 277:6994-7001, [2002]). Chaperones help the secretory protein to fold by binding to exposed hydrophobic regions in the unfolded states and preventing unfavourable interactions. and prolyl-peptidyl *cis-trans* isomerases assist in formation of the proper conformation of the peptide chain adjacent to proline residues.

In some embodiments of the present invention, the host cells are transformed with an expression vector encoding at least one thiol-disulfide oxidoreductase or chaperone. It is not intended that the present invention be limited to any particular thiol-disulfide oxidoreductase or chaperone, as any suitable thiol-disulfide oxidoreductase or chaperone known to those skilled in the art will find use in the present invention.

In some embodiments of the present invention, the fraction of properly folded secretory protein is increased by the addition of chemicals to the growth medium that reduce/oxidize disulfide bonds, and/or alter the general redox potential, and/or chemicals that alter solvent properties thus affecting protein conformation and aggregation. In particularly preferred embodiments, a reagent that reduces disulfide bonds, such as 2-mercaptoethanol, is preferred to increase the fraction of correctly folded protein. However, in other embodiments and depending on the medium used, other disulfide reducing or oxidizing agents (e.g., DTT, TCEP, reduced and oxidized glutathione, cysteine, cystine, cysteamine, thioglycolate, $S_2O_4^{2-}$, $S_2O_3^{2-}$, $S_2O_2^{2-}$, SO_3^{2-} , $S_2O_2^{2-}$, SO_2^{2-} , Cu^{+} , etc.), either used alone or in combination, find use in the present invention. It is contemplated that other adjuvants that alter solvent properties, (e.g., urea, DMSO, TWEEN®-80, etc.), either added to the growth medium alone or preferably in combination with disulfide reducing/oxidizing agents, such as β ME, will also increase the fraction of correctly folded secretory protein and find use in various embodiments of the present invention. In some preferred embodiments, the β ME is used at concentrations ranging from 0.5 to 4 mM, while in other embodiments, the concentrations range from 0.1 mM to 10 mM. Indeed, those of skill in the art know how to select the best growth medium and growth conditions to optimize the effects of the added thiol reducing/oxidizing agents and/or other adjuvants, as well as the concentration of thio reducing/oxidizing agents and/or other adjuvants to use. It is not intended that the present invention be limited to any particular disulfide reducing/oxidizing agent or adjuvant, as any suitable reagents known to those skilled in the art find use in the present invention.

Fermentation Parameters

The present invention relies on fermentation procedures for culturing bacterial species. Fermentation procedures for production of heterologous proteins by bacterial species are well known in the art. Culturing is accomplished in a growth medium comprising an aqueous mineral salts medium, organic growth factors, the carbon and energy source material, molecular oxygen (for aerobic and facultative bacteria), and, of course, a starting inoculum of one or more particular microorganism species to be employed.

In addition to the carbon and energy source, oxygen, assimilable nitrogen, and an inoculum of the microorganism, it is necessary to supply suitable amounts in proper proportions of mineral nutrients to assure proper microorganism growth, maximize the assimilation of the carbon and energy source by the cells in the microbial conversion process, and achieve maximum cellular yields with maximum cell density in the fermentation medium.

Various culture media find use in the present invention, as known to those of skill in the art. However, standard bacterial culture media find use in the present invention. In some preferred media formulations, the media include, in addition to nitrogen, suitable amounts of phosphorus, magnesium, calcium, potassium, sulfur, and sodium, in suitable soluble assemblable ionic and combined forms, and also present preferably should be certain trace elements such as copper, manganese, molybdenum, zinc, iron, boron, and iodine, and others, again in suitable soluble assimilable form, all as known in the art.

In some embodiments, the fermentation reaction involves an aerobic process in which the molecular oxygen needed is supplied by a molecular oxygen-containing gas such as air, oxygen-enriched air, or even substantially pure molecular oxygen, provided to maintain the contents of the fermentation vessel with a suitable oxygen partial pressure effective in assisting the microorganism species to grow in a thriving fashion. In effect, by using an oxygenated hydrocarbon substrate, the oxygen requirement for growth of the microorganism is reduced. Nevertheless, molecular oxygen must be supplied for growth of aerobic and to a lesser extent, facultative organisms.

Although the aeration rate can vary over a considerable range, aeration generally is conducted at a rate which is in the range of about 0.5 to 10, preferably about 0.5 to 7, volumes (at the pressure employed and at 25° C.) of oxygen-containing gas per liquid volume in the fermentor per minute. This amount is based on air of normal oxygen content being supplied to the reactor, and in terms of pure oxygen the respective ranges would be about 0.1 to 1.7, or preferably about 0.1 to 1.3, volumes (at the pressure employed and at 25° C.) of oxygen per liquid volume in the fermentor per minute.

The pressure employed for the microbial conversion process can range widely. Pressures generally are within the range of about 0 to 50 psig, presently preferably about 0 to 30 psig, more preferably at least slightly over atmospheric pressure, as a balance of equipment and operating costs versus oxygen solubility achieved. Greater than atmospheric pressures are advantageous in that such pressures do tend to increase a dissolved oxygen concentration in the aqueous ferment, which in turn can help increase cellular growth rates. At the same time, this is balanced by the fact that high atmospheric pressures do increase equipment and operating costs.

The fermentation temperature can vary somewhat, but for most bacterial species used in the present invention, the temperature generally will be within the range of about 20° C. to 40° C., generally preferably in the range of about 28° C. to 37° C., depending on the strain of microorganism chosen, as known to those skilled in the art.

The microorganisms also require a source of assimilable nitrogen. The source of assimilable nitrogen can be any nitrogen-containing compound or compounds capable of releasing nitrogen in a form suitable for metabolic utilization by the microorganism. While a variety of organic nitrogen source compounds, such as protein hydrolysates, can be employed, usually, cheap nitrogen-containing compounds such as ammonia, ammonium hydroxide, urea, and various ammonium salts such as ammonium phosphate, ammonium sulfate,

ammonium pyrophosphate, ammonium chloride, or various other ammonium compounds can be utilized. Ammonia gas itself is convenient for large scale operations, and can be employed by bubbling through the aqueous ferment (fermentation medium) in suitable amounts. At the same time, such ammonia can also be employed to assist in pH control.

The pH range in the aqueous microbial ferment (fermentation admixture) should be in the exemplary range of about 2.0 to 8.0. However, pH range optima for certain microorganisms are dependent on the media employed to some extent, as well as the particular microorganism, and thus change somewhat with change in media as known to those skilled in the art.

While the average retention time of the fermentation admixture in the fermentor can vary considerably, depending in part on the fermentation temperature and culture employed, as known in the art.

In some embodiments, the fermentation is preferably conducted in such a manner that the carbon-containing substrate can be controlled as a limiting factor, thereby providing good conversion of the carbon-containing substrate to cells and avoiding contamination of the cells with a substantial amount of unconverted substrate. The latter is not a problem with water-soluble substrates, since any remaining traces are readily removed. It may be a problem, however, in the case of non-water-soluble substrates, and require added product-treatment steps such as suitable washing steps. The time needed to reach this limiting substrate level is not critical and may vary with the particular microorganism and fermentation process being conducted. However, it is well known in the art how to determine the carbon source concentration in the fermentation medium and whether or not the desired level of carbon source has been achieved.

Although in some embodiments, the fermentation is conducted as a batch or continuous operation, fed batch operation is generally preferred for ease of control, production of uniform quantities of products, and most economical uses of all equipment.

If desired, part or all of the carbon and energy source material and/or part of the assimilable nitrogen source such as ammonia can be added to the aqueous mineral medium prior to feeding the aqueous mineral medium into the fermentor, indeed, each of the streams introduced into the reactor preferably is controlled at a predetermined rate, or in response to a need determined by monitoring such as concentration of the carbon and energy substrate, pH, dissolved oxygen, oxygen or carbon dioxide in the off-gases from the fermentor, cell density measurable by light transmittancy, or the like. The feed rates of the various materials can be varied so as to obtain as rapid a cell growth rate as possible, consistent with efficient utilization of the carbon and energy source, to obtain a high a yield of microorganism cells relative to substrate charge as possible, but more importantly to obtain the highest production of the desired protein per unit volume.

In either a batch, or the preferred fed batch operation, all equipment, reactor, or fermentation means, vessel or container, piping, attendant circulating or cooling devices, and the like, are initially sterilized, usually by employing steam such as at about 121° C. for at least about 15 minutes. The sterilized reactor then is inoculated with a culture of the selected microorganism in the presence of all the required nutrients, including oxygen, and the carbon-containing substrate. The type of fermentor employed is not critical, though in some embodiments, the 15L Biobalatte (Saint-Germain-en-Laye, France) is preferred.

Protein Separations

In preferred embodiments, once the desired protein is expressed, the secreted protein is recovered. The present invention provides methods of separating a desired protein from its fusion analog. It is specifically contemplated that the methods described herein will find use in the separation of proteinase inhibitor and variants from the fusion analog.

The collection and purification of the desired protein from the fermentation broth can also be achieved using procedures known to those of skill in the art. The fermentation broth will generally contain cellular debris, including cells, various suspended solids and other biomass contaminants, as well as the desired protein product, which are preferably removed from the fermentation broth by means known in the art. Suitable processes for such removal include conventional solid-liquid separation techniques (e.g., centrifugation, filtration, dialysis, microfiltration, rotary vacuum filtration, or other known processes), to produce a cell-free filtrate. In some embodiments, it is preferable to further concentrate the fermentation broth or the cell-free filtrate prior to the purification and/or crystallization process using techniques such as ultrafiltration, evaporation and/or precipitation.

Precipitating the proteinaceous components of the supernatant or filtrate may be accomplished by means of a salt (e.g., ammonium sulfate) or low pH (typically less than 3), followed by purification by a variety of chromatographic procedures (e.g., ion exchange chromatography, affinity chromatography, hydrophobic interaction chromatography, hydrophobic charge induction chromatography etc.) or similar and recognized procedures. It is not intended that the present invention be limited to any particular separation method, as it is contemplated that any method will find use in the present invention.

In certain preferred embodiments, when the expressed desired polypeptide is secreted from the bacterial cells, the polypeptide is purified from the growth media. In preferred embodiments, the expression host cells are removed from the media before purification of the polypeptide (e.g. by centrifugation).

When the expressed recombinant desired polypeptide is not secreted from the host cell, the host cell is preferably disrupted and the polypeptide released into an aqueous "extract" which is the first stage of purification. Preferably, the expression host cells are collected from the media before the cell disruption (e.g. by centrifugation). The cell disruption may be performed by using any suitable means known in the art, such as by lysozyme or beta-glucanase digestion or by forcing the cells through high pressure (See e.g., Scobes, *Protein Purification*, Second edition, Springer-Verlag).

In some embodiments, the addition of six histidine residues (i.e., a "His Tag") to the C-terminus is used as an aid in the purification of the desired protein and its fusion analog. Use of the His tags as a purification aid is well known in the art (See e.g., Hengen, *TIBS* 20:285-286 [1995]). The 6x-tagged proteins are easily purified using Immobilized Metal Ion Affinity Chromatography (IMAC), as known to those skilled in the art.

Purity

For some applications, it is of great importance that the protease inhibitor produced using the present invention be very highly pure (e.g., having a purity of more than 99%). This is particularly true whenever the desired protein is to be used as a therapeutic, but is also necessary for other applications. The methods described herein provide a way of producing substantially pure desired proteins. The desired proteins described herein are useful in pharmaceutical and

personal care compositions. However, it is contemplated that proteins of varying purity levels will be produced using the methods of the present invention and it is not intended that the proteins produced using the present invention be limited to any particular level of purity.

Activation of BBI During Purification

In some embodiments of the present invention, after growth during the purification process, the activity of the protein is increased by the addition of chemicals that reduce/oxidize disulfide bonds and/or alter the general redox potential, and/or chemicals that alter solvent properties thus affecting protein conformation and aggregation. In some particularly preferred embodiments, addition of a reagent that reduces disulfide bonds, such as 2-mercaptoethanol, is used to increase activity of the protein. However, as those skilled in the art appreciate, depending purity and buffer composition, other disulfide reducing or oxidizing agents (e.g., DTT, TCEP) reduced and oxidized glutathione, cysteine, cystine, cysteamine, thioglycolate, $S_2O_2^{2-}$, $S_2O_3^{2-}$, $S_2O_4^{2-}$, SO_4^{2-} , $S_2O_2^{2-}$, Cu^{+2} , protein disulfide isomerases, protein thiol-disulfide oxidoreductases, etc.), either used alone or in combination, find use in the present invention. Other adjuvants that alter solvent properties, (e.g. ethanolamine, DMSO, Tween®-80, arginine, urea, etc.), either added alone or preferably in combination with disulfide reducing/oxidizing agents, such as BME, during the purification process also find use in the present invention by increasing the activity of the protein. In certain preferred embodiments, partially purified protein is diluted in buffer (in some particularly preferred embodiments, a zwitterionic buffer with Tween®-80 at basic pH) and activated with BME and a disulfide oxidizing agent (in alternative preferred embodiments, oxidized glutathione or sodium sulfite).

In addition, it is contemplated that conditions will be screened in order to determine the optimal activation of the protein, if desired. For example, various [BME] concentrations (0.1-10 mM), oxidizing agent concentrations (0 to 1/20 to 20 times the [BME] concentration) pH (7.5-9.5), temperatures (15-40°C), dilutions (1-20 fold), incubation times (12-72 h), aeration (incubations under inert gas to vigorous mixing under oxygen containing gases), buffer types (Tris, CHES, CAPS, Tricine, TAPS, other zwitterionic buffers, etc.), buffer concentrations (0.1-1 M), and the addition of various adjuvants known to alter solvent properties thereby affecting protein conformation and aggregation (e.g., ethanolamine, DMSO, Tween®-80, arginine, urea, etc.) are tested in order to determine the optimum conditions for the expression system used. It is not intended that the present invention be limited to any particular disulfide reducing/oxidizing agent, dilution, temperature, pH, buffer type or composition, or adjuvant, as any suitable reagents known to those skilled in the art find use in the present invention.

EXPERIMENTAL

The present invention is described in further detail in the following Examples which are not in any way intended to limit the scope of the invention as claimed. The attached Figures are meant to be considered as integral parts of the specification and description of the invention. All references cited are herein specifically incorporated by reference for all that is described therein. The following examples are offered to illustrate, but not to limit the claimed invention. In the experimental disclosure which follows, the following abbreviations apply PI (proteinase inhibitor), BBI (Bowman-Birk inhibitor), STI (Soybean Trypsin inhibitor); VEGF and VegF

(vascular endothelial growth factor); ppm (parts per million); M (molar); mM (millimolar); μ M (micromolar); nM (nanomolar); mol (moles); μ mol (millimoles); μ mol (micromoles); nmol (nanomoles); gm (grams); mg (milligrams); ng (micrograms); pg (picograms); L (liters); ml and mL (milliliters); μ l and μ L (microfilters); cm (centimeters); mm (millimeters); μ m (micrometers); nm (nanometers); U (units); V (volts); MW (molecular weight); sec (seconds); min(s) (minute/minutes); hr(s) and hr(s) (hour/hours); °C (degrees Centigrade); QS (quantity sufficient); ND (not done); NA (not applicable); rpm (revolutions per minute); H_2O (water); dH_2O (deionized water); (HCl) (hydrochloric acid); aa (amino acid); bp (base pair); kb (kilobase pair); kD (kilodaltons); cDNA (copy or complimentary DNA); DNA (deoxyribonucleic acid); ssDNA (single stranded DNA); dsDNA (double stranded DNA); dNTP (deoxyribonucleotide triphosphate); RNA (ribonucleic acid); Mg²⁺ (magnesium chloride); NaCl (sodium chloride); w/v (weight to volume); v/v (volume to volume); g (gravity); OD (optical density); Dulbecco's phosphate buffered solution (DPBS); SOC (2% Bacto-Tryptone, 0.5% Bacto Yeast Extract, 10 mM NaCl, 2.5 mM KCl); Terrific Broth (TB: 12 g/l Bacto Tryptone, 24 g/l glycerol, 2.31 g/l KH_2PO_4 and 12.54 g/l K_2HPO_4); OD₂₆₀ (optical density at 260 nm); OD₄₀₅ (optical density at 400 nm); A_{405} (absorbance at 405 nm); Vmax (the maximum initial velocity of an enzyme catalyzed reaction); PAGE (polyacrylamide gel electrophoresis); PBS (phosphate buffered saline [150 mM NaCl, 10 mM sodium phosphate buffer, pH 7.2]); PBST (PBST+0.25% Tween® 80); PEG (polyethylene glycol); PCR (polymerase chain reaction); RT-PCR (reverse transcription PCR); SDS (sodium dodecyl sulfate); Tris (tris(hydroxymethyl)aminomethane); HEPES (N-[2-Hydroxyethyl]piperazine-N-[2-ethanesulfonic acid]); HBS (HFFES buffered saline); SDS (sodium dodecylsulfate); bME, BME and β ME (beta-mercaptoethanol or 2-mercaptoethanol); Tris-HCl (tris(hydroxymethyl)aminomethane-hydrochloride); Tricine (N-[tris(hydroxymethyl)-methyl]-glycine); CHES (2-(N-cyclo-hexylamino) ethane sulfonic acid); TAPS (3-[(tris(hydroxymethyl)-methyl)-amino]-propane-sulfonic acid); CAPS (3-(cyclo-hexylamino)-propano-sulfonic acid); DMSO (dimethyl sulfoxide); DTT (1,4-dithio-DI-threitol); Glut and GSH (reduced glutathione); GSSG (oxidized glutathione); TCEP (Tris[2-carboxyethyl] phosphine); Ci (Curies); mCi (milliCuries); μ Ci (microCuries); TLC (thin layer chromatography); Ts (tosyl); Bn (benzyl); Ph (phenyl); Ms (mesyl); Et (ethyl); Me (methyl); Taq (*Thermus aquaticus* DNA polymerase); Klenow (DNA polymerase I large (Klenow) fragment); rpm (revolutions per minute); EGTA (ethylene glycol-bis(β-aminoethyl ether) N,N,N',N'-tetraacetic acid); EDTA (ethylenediaminetetraacetic acid); bla (β -lactamase or ampicillin-resistance gene); GE Healthcare (GE Healthcare, Chalfont St. Giles, United Kingdom); DNA2.0 (DNA2.0, Menlo Park, Calif.); OXOID (Oxoid, Basingstoke, Hampshire, UK); Megazyme (Megazyme International Ireland Ltd., Bray Business Park, Bray, Co. Wicklow, Ireland); Coming (Corning Life Sciences, Coming, N.Y.); (NFN) (NEN Life Science Products, Boston, Mass.); Pharma/AS (Pharma/AS, Oslo, Norway); Dynal (Dynal, Oslo, Norway); Bio-Synthesis (Bio-Synthesis, Lewisville, Tex.); ATCC (American Type Culture Collection, Rockville, Md.); Gibco/BRL (Gibco/BRL, Grand Island, N.Y.); Sigma (Sigma Chemical Co., St. Louis, Mo.); Pharmacia (Pharmacia Biotech, Piscataway, N.J.); NCB (National Center for Biotechnology Information); Applied Biosystems (Applied Biosystems, Foster City, Calif.); Clontech (CLONTECH Laboratories, Palo Alto, Calif.); Operon Technologies (Operon Technologies, Inc., Alameda, Calif.); MWG Biotech

(MWG Biotech, High Point, N.C.); Oligos Etc (Oligos Etc. Inc., Wilsonville, Ore.); Bachem (Bachem Bioscience, Inc., King of Prussia, Pa.); DiFCO (DiFCO Laboratories, Detroit, Mich.); Mediatech (Mediatech, Herndon, Va.; Santa Cruz (Santa Cruz Biotechnology, Inc., Santa Cruz, Calif.); Bio-Veris (BioVeris Corp., Gaithersburg, Md.); OxoID (OxoID Inc., Ogdensburg, N.Y.); Worthington (Worthington Biochemical Corp., Freehold, N.J.); Gibco BRL or Gibco BRL (Life Technologies, Inc., Gaithersburg, Md.); Millipore (Millipore, Billerica, Mass.); Bio-Rad (Bio-Rad, Hercules, Calif.); Invitrogen (Invitrogen Corp., San Diego, Calif.); NFB (New England Biolabs, Beverly, Mass.); Sigma (Sigma Chemical Co., St. Louis, Mo.); Pierce (Pierce Biotechnology, Rockford, Ill.); Takara (Takara Bio Inc. Otsu, Japan); Roche (Hoffmann-La Roche, Basel, Switzerland); EM Science (EM Science, Gibbstown, N.J.); Qiagen (Qiagen, Inc., Valencia, Calif.); Biodesign (Biodesign Int'l., Saco, Me.); Aptagen (Aptagen Inc., Herndon, Va.); Molecular Devices (Molecular Devices Corp., Sunnyvale, Calif.); R&D Systems (R&D Systems, Minneapolis, Minn.); Stratagene (Stratagene Cloning Systems, La Jolla, Calif.); Marsh (Marsh Biosciences, Rochester, N.Y.); Bio-Tek (Bio-Tek Instruments, Winooski, VT); Biacore (Biacore Inc., Piscataway, N.J.); PeproTech (PeproTech, Rocky Hill, N.J.); SynPep (SynPep, Dublin, Calif.); and Microsoft (Microsoft, Inc., Redmond, Wash.).

Example 1

Production of BCE103-BBI Fusion Proteins in *B. subtilis*

In this Example, experiments conducted to produce BCE103-BBI fusion proteins in *B. subtilis* are described. The DNA sequence of the synthetic gene (Operon Technologies) coding for the pro-BBI protein with a C-terminal hexa-histidine tag used in these experiments is:

(SEQ ID NO:10)
 AACCTGCGCTGCTAAQCTTGGCTGCTTATGAAATCAGGACCATCGAGCA
 CAGCAATGAGCATGAGGAGCTCTAAACCCCTGTCGGATCATGGCAGATGTA
 CGAAATCAATCTCCACAGTGTGGTGTCCGATATGCGCTGAGAC
 TGTCTAGTGCGATCGAAAGCTGTATCTGGCCCTGAGTTATCGAGCTA
 ATGTTTTGGCTCGACATCACGAGCTTCGCTGATGAGCCATGTAACCAA
 GCGAGGACGATAAAGGAGAACCATCACCCATCACCAT

The protein sequence of pro-BBI with a C-terminal hexa-histidine tag used in these experiments is:

(SEQ ID NO:11)
 NLRLSKLGLLMKSDHQHSHNDESSKPCCDQCACTESNPPQCRCSDMRLNS
 CHSACKSCSICALSYPAQCFCV11TDPCYEPCKPSEDDKEHHHHHHHH

The portion of the DNA sequence of the synthetic gene that codes for the major mature form of BBI is:

(SEQ ID NO:12)
 GAGCATGAGGAGCTCTAAACCCCTGTCGGATCATGGCAGATGAGAACAT
 AAATCCCTCACAGTGTGGTGTCCGATATGCGCTGATGAGCTGCTGATA
 GTGCAATGCAAAAGCTGTATCTGGCCCTGAGTTATCCAGCTAATGTTT

-continued
 TGGCTGACATCACGGACTCTGCTATGAGGCATGTAACCAAGCGAGGA
 CGATAAAAGGAGAC

The protein sequence of the major mature form of BBI coded by the above synthetic gene is:

(SEQ ID NO:13)
 DDESSKPCCDQCACTESNPPQCRCSDMRLNSCHSACKSCSICALSYPAQCFC
 CVDITDPCYEPCKPSEDDKEEN

The PCR primers used to amplify the BBI gene for fusion 15 to the BCE103 cellulase expression cassette in the pJ103 vector were:

(SEQ ID NO:14)
 20 BBIfusion_FW: 5' CAGCACGGATCCAGCATGAGAGCTCT
 AAACCC 3'
 BBIHindIII_RV: 5' CTGCGAGAAGCTTTAAATTAATTTAAACCGGA

(SEQ ID NO:15)
 25 TTTCTTCAGGAAATCCGTCCTCTGTAACTT
 TTAGTTCTCTTTATGTCCTCGC 3'

(SEQ ID NO:16)
 30 BBIHIS-HindIII_RV: 5' CTGCGAGAAGCTTTAAATTAATTTAAACCGGA
 TTTCTTCAGGAAATCCGTCCTCTGTAACTT
 TTATGGTATGGTGTATGGTCTC 3'

The sequence of the aprE-BCE103-BBI-HisTag expression cassette (EcoRI-HindIII) that was cloned into the pJ103 expression vector is provided in FIG. 1. A schematic map of the pJ1M103BBIHis expression vector is provided in FIG. 2.

The alkaline cellulase (BCE103) gene (See, van Soligen, U.S. Pat. No. 6,063,611, hereby incorporated by reference) fused to the *B. subtilis* aprE promoter and signal sequence, was cloned from pUCAPR103 (Shaw et al., J. Mol. Biol., 320:303-309 [2002]) as an EcoRI-BamHI fragment (i.e., a fragment that carries the coding sequence of the BCE103

catalytic domain and first cellulose binding domain linker only) into pJM103 (Pergo, "Integrational vectors for genetic manipulation in *Bacillus subtilis*" In, *Bacillus subtilis* and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics, Sonenshein, Hoch, and Losick (eds.), American Society for Microbiology, Washington D.C., pp. 615-624 [1993]). A gene encoding the soybean Bowman-Birk protease inhibitor (BBI) (Swiss-Prot Accession # P0105; See, Odani and Ikenaka, J. Biochem., 71: 839-848 [1972]) with a C-terminal hexa-histidine tag (His-Tag) was synthesized by Operon Technologies (See, DNA sequence above). The BBI gene was amplified by PCR with primers (all primers were synthesized by MWG Biotech, Oligos Etc., or Operon Technologies) that generated a 5' BamHI site in the correct reading frame with the BCE103 gene, and at the 3' end introduced a strong transcriptional terminator (LAI), from the *Bacillus licheniformis* α -amylase gene, after the end of the BBI gene with a 3' HindIII site for cloning into the pJM103 vector.

PCR fragments with or without a C-terminal His-Tag were generated with the primers BBIfusion_FW (SEQ ID NO:14) and BBIHIS-HindIII_RV (SEQ ID NO:16), or BBIfusion_FW (SEQ ID NO:14) and BBI-HindIII_RV (SEQ ID NO:15), respectively, using the synthetic BBI gene as a tem-

plate. Unless indicated otherwise, PCR reactions were typically performed on a thermocycler for 30 cycles with High Fidelity Platinum Taq polymerase (Invitrogen) according to the instructions of the supplier (with an annealing temperature of 55°C). The PCR fragments were cloned as BamHI-HindIII fragments into pM103 carrying the aprE-BCE 103 expression cassette. The correct gene sequence was verified by DNA sequencing.

Thus, as shown in FIG. 1, the N-terminus of the mature coding region of the BBI gene (with or without the His-Tag) was fused in frame to the C-terminal coding region of the first CBD (cellulose binding domain) linker sequence coded by the BCE103 cellulase gene. Thereby, the two CBD's of BCE103 (Shaw et al., supra) are replaced by BBI in the final expression vectors, pM103BBI or pM103BBH1s (See, FIG. 2). The aprE promoter controls the expression of the BCE103-BBI gene fusions (See, Ferrari et al., J. Bact., 170: 289-295 [1988]; and Henner et al., J. Bact., 170: 296-300 [1988]).

Competent *Bacillus subtilis* cells, BG3934comK, (degU⁺32, oppA, ΔpolI⁺E3501, ΔaprE, ΔaprP, ΔispA, Δapr, amyE:zLRPxy1/ΔcomK-phle) were transformed with the expression plasmids, pM103BBI or pM103BBH1s. The bacteria were made competent by the induction of the comK gene under control of a xylose inducible promoter (Hahn et al., Mol. Microbiol., 21:763-775 [1996]). The transformants were selected on Luria Broth agar (LA) plates containing 5 µg/ml chloramphenicol. To increase the expression by gene amplification, colonies were streaked and grown several times on LA plates with 25 µg/ml chloramphenicol until the growth rate with the antibiotic was similar to growth rate in the absence of chloramphenicol. The BCE103-BBI fusion protein was produced by growth in shake flasks at 37°C in TSB medium (Tryptone Soya Broth from OXOID, 30 g/L) or in MBD medium, a MOPS based defined medium. MBD medium was made essentially as described (Neidleman et al., J. Bacteriol., 119: 736-747 [1974]), except NH₄Cl, 3 mM FeSO₄, and CaCl₂ were left out of the base medium, 3 mM K₂HPO₄ was used, and the base medium was supplemented with 60 mM urea, 75 g/L glucose, and 1% soytoye. Also, the microcuitrients were made up as a 100x stock containing in one liter, 400 mg FeSO₄ 7H₂O, 100 mg MgSO₄·H₂O, 100 mg ZnSO₄·7H₂O, 50 mg CuCl₂·2H₂O, 100 mg CoCl₂·6H₂O, 100 mg NaMoO₄·2H₂O, 100 mg Na₂B₄O₇·10H₂O, 10 ml of 1M CaCl₂, and 10 ml of 0.5 M sodium citrate.

BCE103-BBI fusion protein could be easily visualized in samples from cell free supernatants (after 24 h of growth in TSB medium or 4 h in MBD medium) as the major protein band on SDS-PAGE gels (10% NuPAGE in MES buffer, run as described by the manufacturer, Invitrogen) running at ~44 kDa by using standard protein stains (e.g. GelCode Blue Stain Reagent; Pierce). The identity of the BCE103-BBI fusion protein was verified by immunoblots of SDS-PAGE gels using the protocols supplied by the manufacturer (BM Chromogenic Western Blotting Kit; Roche Applied Science using an anti-HisTag antibody or an anti-BCE103 cellulase polyclonal antibody for detection).

To determine the BCE103 activity, cellulase degradation was assessed qualitatively on LA cellulase indicator plates (with 1% carboxymethylcellulose stained with 0.2% Congo Red, or with 0.5% azo-CM-cellulose, Megazyme), or quantitatively by a direct assay in Assay Buffer (100 mM Tris pH 8.6, 0.005% Tween-80) on the culture broth using a the synthetic substrate, 4-nitrophenyl β-D-celluloside (Sigma), using methods known in the art (See e.g., van Tilbeurgh et al., Meth. Enzymol., 160:45-59 [1988]).

TrpS inhibitory assays were performed in Assay Buffer to determine the BBI activity. Specifically, a standard curve was generated by making eleven 1:1 serial dilutions (100 µL

BBI+100 µL Assay Buffer) of a 2 µg/ml standard BBI solution. The BBI standard was purified from a 1 mg/ml Trypsin-Chymotrypsin Inhibitor (Sigma Cat. #T-9777) solution in 20 mM MES pH 6.0 using a hydrophobic interaction column (POROS HP2, Phenyl column, Applied Biosystems). The column was equilibrated with 20 mM MES pH 6.0, loaded with 5 mg of the inhibitor, washed with the equilibration buffer, and then the BBI was eluted with water. Unknown BBI samples to be tested in the inhibitory assay were diluted as necessary, so that two or more data points would fall within the standard curve (usually 1:10, 1:100, 1:200, 1:1000, 1:2000 sample dilutions were tested and then the dilutions fine tuned if necessary). Each diluted BBI standard or sample, 20 µL, was added to 80 µL of 50 ng/ml bovine pancreatic trypsin (Worthington) (made by diluting a stock 1 mg/ml trypsin solution into Assay Buffer). For convenience, the standards and samples were arrayed in 96 well microtiter plates. The reactions were mixed and incubated 15 min at 25°C. After the incubation, 100 µL of the 0.5 mg/ml trypsin substrate (diluted in Assay Buffer from a 100 mg/ml solution in DMSO), Suc-AAPR-pNA (succinyl-Ala-Ala-Pro-Arg-p-nitroanilide, Bachem), was added, mixed and the OD (A₄₀₅) was monitored for 15 min, with 1 time point recorded every 12 sec using a Spectra Max 250 (Molecular Devices). The data points were used to determine the Vmax for each reaction. The standard curve was generated by plotting Vmax versus BBI concentration and was fitted to a four-parameter curve. All data fitting was done using software supplied by the manufacturer (Molecular Devices). The BBI concentration of the unknown samples was calculated from the standard curve. Alternatively, the BBI activity was measured using the same protocol but by determining bovine pancreatic chymotrypsin (Worthington) inhibition (chymotrypsin was used at the same concentration as trypsin and chymotrypsin activity was measured by adding 100 µL of a 0.4 mg/ml chymotrypsin substrate, succinyl-Ala-Ala-Pro-Phe-p-nitroanilide, Bachem).

Titers from shake flask runs (500 ml MBD medium in 2.8 L Fernbach 6 baffled flasks, 37°C, 225 rpm, harvested 60 h after growth) typically ranged from 0.4-0.9 mg/ml BCE activity and 40-150 µg/ml BBI trypsin inhibitory activity. However, it is contemplated that titers likely could be improved further by optimizing the bacterial strain, culture medium and growth conditions (aeration, temperature, time of harvest, etc.).

In addition to the BCE103 fusion to wild-type BBI, fusion proteins to BBI variants and fusion proteins with various linkers between BCE103 and BBI were produced using the methods outlined above, as described in the following Examples. In addition, fusion proteins were also produced when the BBI was fused to the 2nd CBD linker (BCE-cbd2-BBI; See, Example 4) making it possible to use the 1st CBD to aid in the purification process.

Example 2

Production of Peptides Substituted into the BBI Reactive Site Loops as BCE103-BBI Fusion Proteins

In this Example, experiments conducted to produce peptides substituted into the BBI reactive site loops as BCE103-BBI fusion proteins are described. The primers, as well as other sequences used in the various steps of these experiments are provided below. The sequence of 12BBIekR1 from the BCE103 fusion site (at the BamHI) to the end of the gene is provided in FIG. 3. The CK37281 peptide sequences (ACYNLYGWTC (SEQ ID NO:9) are inserted into both the trypsin and chymotrypsin inhibitory loops.

The primers used to introduce an EcoRI site in the BBI gene using QuikChange® site-directed mutagenesis (Stratagene) were:

-continued

BowBeco-P
5'-GATATGCGCTGGAATTCCTGCTATGTCAT (SEQ ID NO:17)

BowBeco-R
5'-ATGCACTATGACAGGAATTTCAGACCATATC (SEQ ID NO:18)

The sequences of the DNA oligonucleotides that were annealed and cloned in the BBI gene (Sacl-EcoRI) to replace the trypsin inhibitory loop with the Vegf' binding peptide CK37281 were:

1BBck81+ (SEQ ID NO:19)
5'-CTAACCTGTGCGATCAATGGCGATTTATAATTGTATGGGTG

ACTTGTCGCTGAGCGATATGCGCTCG

1BBck81- (SEQ ID NO:20)
5'-AATTCAAGCGCATATCCTGCGACGAGTCGACCCATACRAAATTA

TAACATGCCATGTATGCCAACAGGGTTTAGAGCTT

The sequences of the DNA oligonucleotides that were annealed and cloned in the BBI gene (EcoRI-SalI) to replace the chymotrypsin inhibitory loop with the Vegf' binding peptide CK37281 were:

2BBck81+ (SEQ ID NO:21)
5'-AATTCCGTATAGTCGCTGCAAAGGTGGCATGTTATAACCTGTA

CGGGTGGACCTGTTTTCG

2BBck81- (SEQ ID NO:22)
5'-TCGACCAAAAAAAGGTCACCGGTACAGGTATAACATGGCGAGCT

TTTCAAGGGACTATGACCGG

The DNA sequences of the oligonucleotide pairs used to make cassettes to introduce peptides into the trypsin (Sacl and EcoRI restriction sites) or chymotrypsin (EcoRI and SalI restriction sites) reactive site loops of the synthetic BBI gene are provided below. These peptide coding sequences were then moved into the p2JM103BBI expression vector as Sacl-SalI fragments.

Comstatin (1st loop) (SEQ ID NO:23)
CTAACCTGTGCGATCATGGCGATGTGTTCAAGGACTGGGGTCAC

CACCGTTGCGCTGCGAGCGATATGCGCTCG
and

(SEQ ID NO:24)
AATTTCAAGACGCAATATGCGCTGCGAGCGACAACGGTTTATAGAGCTT

GAACAAACATGCAATTGATGCCAACAGGGTTTATAGAGCTT

Comstatin (2nd loop) (SEQ ID NO:25)
CAAAAGCTGATCTCGCTTGTTCAGACTGGGGTCACACCGTTG

TTTTTCG
and

(SEQ ID NO:26)
TCGACGCAAAACAAACGGTGGTGAACCCAGTCTGAAACACAGATAC

AGCTTTTCGATG

C2c (1st loop) (SEQ ID NO:27)
CTAACCTGTGCGATCAATGCGCTGCGATGTTAAACCCGATCCA

GTGTCGCTGCGAGCGATATGCGCTCG

and

5'-ATTCAAGCGCATATGCGTCGAGCGACACTGGATCGGGATTTT (SEQ ID NO:28)

ACGACACAGCGTCAATTGATGATCACAACAGGGTTTAGAGCTT

10 C3c (1st loop) (SEQ ID NO:29)
CTAACACCCTGTTGCGATCAATGCGGTGTTGCTGCGTCTAACCTGGAGCGAA

TGTCGCTGCGAGCGATATGCGCTCG

and

15 (SEQ ID NO:30)
ATTCAAGCGCATATGCGTCGAGCGACATTGCGTCCAGGGTTAGAGCTT

ACAAACCGATTGATGCCAACAGGGTTTAGAGCTT

20 C4c (1st loop) (SEQ ID NO:31)
CTAACACCCTGTTGCGATCAATGCGGTGTTGCTGCGTCTAACCTGGAGCTG

TGTCGCTGCGAGCGATATGCGCTCG

and

25 (SEQ ID NO:32)
ATTCAAGCGCATATGCGTCGAGCGACACAGGGATCGGGAGCGACCTGA

CAACCGCATTGATGCCAACAGGGTTAGAGCTT

30 C5c (1st loop) (SEQ ID NO:33)
CTAACACCCTGTTGCGATCAATGCGGTGTTGCTGCGTCTAACCTGGAGAAAC

TGTCGCTGCGAGCGATATGCGCTCG

and

35 (SEQ ID NO:34)
ATTCAAGCGCATATGCGTCGAGGGACAGGGTTTATGCGAGCGACCA

CACTGGCATTGATGCCAACAGGGTTAGAGCTT

Xa1 (2nd loop) (SEQ ID NO:35)
40 ATTCTGCTGTCATAGTGCCTGCAAAAGCTGTATCTGGGCCGTTAGTTGCC

AGCTCAATGTTTTCG
and

45 (SEQ ID NO:36)
TCGACGCAAAACATGAGCTGGCAACTACGGGCACAGATAACGGTTT

GCAGGCACTATGACAGG

hscC1 (1st loop) (SEQ ID NO:37)
50 CTAACACCCTGTTGCGATCAATGCGTCGAGCGACTGTTGAGCTACACCCCTCCACAG

TGTCGCTGCGAGCGATATGCGCTCG

and

55 (SEQ ID NO:38)
ATTCAAGCGCATATGCGTCGAGGGACACTGTTGAGGGTTGAGCTA

CACTGGCATTGATGCCAACAGGGTTAGAGCTT

The DNA sequences of oligonucleotide primer pairs used to introduce peptide sequences into the trypsin or chymotrypsin reactive site loops using a QuikChange® II XL site-directed mutagenesis kit (Stratagene) are provided below.

The reactions were performed as outlined by the manufacturer and described in this Example. Twenty cycles were performed with extensions of 6 minutes at 68° C., denaturations of 50 s at 95° C., and annealings at 55° C. for 50 s. After the cycles, a final extension was performed at 68° C. for 20 minutes.

1A (2nd loop)
 CTG TAT CTC A A A C G C T C A A A A T C T C G T G G C T G T T T G C G T C G A C A T C A C
 and
 CG C A A A A C A G C C A C G A G A T T T G N G C G T T G C A G A T A C A G C T T T G C A T G
 (SEQ ID NO:39)

2B (2nd loop)
 CTG TAT C T G C T G G T A T A C C A A A T G A C A C A T G T T T G C G T C G A C A T C A C
 and
 CG C A A A A C A T G T G T G A T T G A T G G C A G A T A C A G C T T T G C A T G
 (SEQ ID NO:40)

4A (2nd loop)
 CTG TAT C T G C C A T C A C T T G G C C G A A T T C A T G T T T G C G T C G A C A T C A C
 and
 CG C A A A A C A T G A T T G C G G C C A A G T G T G A T G G C A G A T A C A G C T T T G C A T G
 (SEQ ID NO:41)

5A (2nd loop)
 CTG TAT C T G C C A T C C G T G G G C A C C G T A T T C T G T T T G C G T C G A C A T C A C
 and
 CG C A A A A C A G A A T A C C G G T G C C C A C G G A T G G C A G A T A C A G C T T T G C A T G
 (SEQ ID NO:42)

6-1A (2nd loop)
 CTG TAT C T G C A A T C T C A T T A T C T C A A C A G C T T T T G C G T C G A C A T C A C
 and
 CG C A A A A C A G T G T G A A G A T A A T G P A G A T T G C A G A T A C A G C T T T G C A T G
 (SEQ ID NO:43)

7A (2nd loop)
 CTG TAT C T G C A C A C C O T C T C T T T A T C G C C C G T G T T T T G C G T C G A C A T C A C
 and
 CG C A A A A C A C G G C G A T A A R G A G A C G G T G T G C A G A T A C A G C T T T G C A T G
 (SEQ ID NO:44)

8B (2nd loop)
 CTG TAT C T G C C T T A C A G A T C A A T C T A A A C C G T G T T T T G C G T C G A C A T C A C
 and
 CG C A A A A A C A C G G T T G A T G A T G T G T G A T G G C A G A T A C A G C T T T G C A T G
 (SEQ ID NO:45)

9A (2nd loop)
 CTG TAT C T G C G T A C A C A T C A A T G G G C A T G T G T T T T G C G T C G A C A T C A C
 and
 CG C A A A A A C A C T G C C C A T T G A T G T G T A C C G A G A T A C A G C T T T G C A T G
 (SEQ ID NO:46)

10B (2nd loop)
 CTG TAT C T G C C G C G C A T C C C G T A T G A T G T T T T T G C G T C G A C A T C A C
 and
 CG C A A A A A C A C C A C A T C A C G G T G A T G C G C G G C A G A T A C A G C T T T G C A T G
 (SEQ ID NO:47)

11-1A (2nd loop)
 CTG TAT C T G C T A C A C A C A A A A A T T C C G C A T G T T T T G C G T C G A C A T C A C
 and
 CG C A A A A A C A T T G C G G A A T T T T T G O T G T G A C G A G A T A C A G C T T T G C A T G
 (SEQ ID NO:48)

12B (2nd loop)
 CTG TAT C T G C A C A C A T T T G C T C T G C A C A T G T T T T G C G T C G A C A T C A C
 and
 CG C A A A A A C A T G T G C A G G C G A A T T G T G C G A G A T A C A G C T T T G C A T G
 (SEQ ID NO:49)

13A (2nd loop)
 CTG TAT C T G C C C G G A T C A T G T C C G C A T C T T G T T T T G C G T C G A C A T C A C
 and
 CG C A A A A A C A C A A G A T O C G G A A C N T Q A T C C G G C A G A T A C A G C T T T G C A T G
 (SEQ ID NO:50)

15-1A (2nd loop)
 CTG TAT C T G C T C A G G C T T T C C O C T T C A T G T T T T G C G T C G A C A T C A C
 and
 CG C A A A A A C A T G T A G A A A G C G G R A A G C C G G A G C G T G A G C A G A T A C A G C T T T G C A T G
 (SEQ ID NO:51)

1A6 (1st loop)
 T C A A T G C G C A T G T G A G G A G A T C T G G C T A T G O T C T G C C G G T G C C C G A T A T G C G T C
 (SEQ ID NO:52)

-continued

and

CGGAAACCCGGCAAAGCATAGTCCAGATCTCTTCACATGCGCATTGATCGAACAGG (SEQ ID NO:66)
 1A6 (2nd loop)
 CAAAGCTGTGCTTGTGAGAGATCTGGAGATCTGCTTGTCTTTCGATCGACATCACGG (SEQ ID NO:67)
 and
 ACGCCAAAGGAAACATGAGTCTCCAGATCTCTTCACAGCAGCTTTCATGCGACTATG (SEQ ID NO:68)
 1C2 (1st loop)
 TCAATCGCAGATGTTGGGCCCTTACTGTGCAAAACATGCGCATTGATCGAACAGG (SEQ ID NO:69)
 and
 CGGAAACCCGGCATGTTTGAGACTAAAGGGCCCAAACAGCACAGCTTTCATGCG (SEQ ID NO:70)
 1C2 (2nd loop)
 CAAAGCTGTGCTTGTGAGTCTCCAGTACTGTGCAAAACATGCGCATTGATCGAACAGG (SEQ ID NO:71)
 and
 ACGCCAAAGGATGTTTGAGACTAAAGGGCCCAAACAGCACAGCTTTCATGCGACTATG (SEQ ID NO:72)
 2E2 (1st loop)
 TCAATCGCAGATCTTCAAGTACTGTGGACTACATGCTTTCGCGATATGCGTC (SEQ ID NO:73)
 and
 CGGAAACACCGGATGAGTCTCCACAGTACTGTAAAGACATGCGCATTGATCGAACAGG (SEQ ID NO:74)
 2E2 (2nd loop)
 CAAAGCTGTGCTTGTCTTACAGTACTGTGGACTACATGCTTTCGCGATCGACATCACGG (SEQ ID NO:75)
 and
 ACGCCAAAGGATGAGTCTCCACAGTACTGTAAAGACAGCACAGCTTTCATGCGACTATG (SEQ ID NO:76)
 2E5 (1st loop)
 TCAATCGCAGATCTCTTGGAAAGAGATCGCTTGGCGGTGTTCCGATATGCGTC (SEQ ID NO:77)
 and
 CGGAAACACCGGCAAGGAGATGCTGTTCCAAAGAGATACATGCGCATTGATCGAACAGG (SEQ ID NO:78)
 2E5 (2nd loop)
 CAAAGCTGTGCTTGTACTCTTGGAAATGCTCTCTTCTCTTTCGATCGACATGACGG (SEQ ID NO:79)
 and
 ACGCCAAAGGAAAGAGATGCTGTTCCAAAGAGTACAACAGCACAGCTTTCATGCGACTATG (SEQ ID NO:80)
 FGFn1 (1st loop)
 TCAATCGCAGATGCAAAACATGCTTCTACTCTCTTGGCGGTGTTCCGATATGCGTC (SEQ ID NO:81)
 and
 CGGAAACACCGGCAAGGAGATGAGATGTTGTACATGCGCATTGATCGAACAGG (SEQ ID NO:82)
 FGFn1 (2nd loop)
 CAAAGCTGTGCTTGTGACAAACATGCGATCTACTCTCTTGGCGTGCACATCACGG (SEQ ID NO:83)
 and
 ACGCCAAAGGAAAGAGATGAGATGTTGTGCAAGACAGCTTTCATGCGACTATG (SEQ ID NO:84)
 FGFr1 (1st loop)
 TCAATCGCAGATGCAAAACATGCGTACTCTCTTGGCGGTGTTCCGATATGCGTC (SEQ ID NO:85)
 and
 CGGAAACACCGGCAAGGAGATGAGATGTTGTACATGCGCATTGATCGAACAGG (SEQ ID NO:86)
 FGFr1 (2nd loop)
 CAAAGCTGTGCTTGTGACAAACATGCGTACTCTCTTGGCGTGCACATCACGG (SEQ ID NO:87)
 and
 ACGCCAAAGGAAAGAGATGAGATGTTGTGCAAGACAGCTTTCATGCGACTATG (SEQ ID NO:88)
 FGPh1 (1st loop)
 TCAATCGCAGATGCAACCTGCAAGACAACTGAAACATGCGCGGTGTTCCGATATGCGTC (SEQ ID NO:89)
 and
 CGGAAACACCGGCAAGGAGATGAGATGTTGTGCAAGACAGCTTTCATGCGCATTGATCGAACAGG (SEQ ID NO:90)

-continued

PGP1 (2nd loop)
 CAAAGAGCTGTCCTTGCCACCTGCAGACAACATGAAACATGTTTGCCTGCACATCACCG
 and
 (SEQ ID NO: 91)

ACGCCAAAAACATTTTCAAGTTGCTGCAAGGTTGCAAGCACAAGCTTTGCATGCACATAG
 (SEQ ID NO: 92)

PGFg7 (1st loop)
 TCAATGGCGATGTTGGCTACTTCATCCCCATCGATTGGCGGTTGGCTGATATGCGT
 and
 (SEQ ID NO: 93)

CGGAAACACCGGCCAATTCGATGGATGAAGTAGGCCACATGCCTGGATTGATCACAACAGG
 (SEQ ID NO: 94)

PGFg7 (2nd loop)
 CAAJAGCTGTCCTTGCGGCTACTTCATCCCCATCGATTGGCGGTTGGCTGACATCACCG
 and
 (SEQ ID NO: 95)

ACGCCAAAAACAAATCGATGGGATGAAGTAGGCCCAAGCACAAGCTTTGCATGCACATAG
 (SEQ ID NO: 96)

MM005 (1st loop)
 TCAATGGCGATGTTGGCTACTTCATCCCCATCGATTGGCGGTTGGCTGATATGCGT
 and
 (SEQ ID NO: 97)

CGGAAACACCGGCCATTTGTTAGCAAGGATACTGTAACATGCCTGGATTGATGCCAACAGG
 (SEQ ID NO: 98)

MM005 (2nd loop)
 CAAJAGCTGTCCTTGCTTACGATACCTCTGCTAACAAATGTTTGCCTGACATCACCG
 and
 (SEQ ID NO: 99)

ACGCCAAAAACAAATTTGTTAGCAAGGATACTGTAACATGCCTGGATTGATGCCAACAGG
 (SEQ ID NO: 100)

MM007 (1st loop)
 CGGATCAATGGCGCTGCAGAACTCAACCATATCCTTATGCTGGTTGGCTGATATGCGT
 and
 (SEQ ID NO: 101)

GGAAACACCGGCCATTAAGGATATGGTTGAGTTGCTGAGGGCGCATTCGCAACACGGTT
 (SEQ ID NO: 102)

MM007 (2nd loop)
 CAAAGCTGTCCTGCTCAGAACACACCTTACCCACTTACCTGTTTGCCTGACATCACCG
 and
 (SEQ ID NO: 103)

ACGCCAAAAACAAAGTGGGTAAGGTTGTTCTGCAGGCACAGCTTTGCATGCACATAG
 (SEQ ID NO: 104)

MM009 (2nd loop)
 CAAAGCTGTCCTGTTAACACCTACTCTAACACTGTTTGCCTGACATCACCG
 and
 (SEQ ID NO: 105)

ACGCCAAAAACAGTTAGAGTAAAGTTACAGGCAGGGCACAGCTTTGCATGCACATAG
 (SEQ ID NO: 106)

MM010 (1st loop)
 TCAATGGCGATGCTCTCCAACTCACTCTAACACTGTTGGCTGTTGGCTGATATGCGT
 and
 (SEQ ID NO: 107)

CGGAAACACCGCACAGTTAGAGTGGAGAGGCGCAGGTTGCTGCAACAGG
 (SEQ ID NO: 108)

MM010 (2nd loop)
 CAAAGCTGTCCTGCGCCTTCCTACACACTCTAACACTGTTTGCCTGACATCACCG
 and
 (SEQ ID NO: 109)

ACGCCAAAAACAGTTAGAGTGGAGAGGCGCAGGACAGCTTTGCATGCACATAG
 (SEQ ID NO: 110)

MM017 (2nd loop)
 CAAAGCTGTCCTGCGCCTTTAGGCTTGGCCACCTGTTTGCCTGACATCACCG
 and
 (SEQ ID NO: 111)

ACGCCAAAAACAGGTGTCAGGCAAGGCTAAAGGGCAGGGCACAGCTTTGCATGCACATAG
 (SEQ ID NO: 112)

PGPpol (2nd loop)
 AAAGCTGATCTGCTGGAAAGATGATTCTACACCTTGTGTTGCCTGACATCACCG
 and
 (SEQ ID NO: 113)

ACGCCAAAAACAGGTGTCAGGCAAGGCTAAAGGGCAGGGCACAGCTTTGCATGCACATAG
 (SEQ ID NO: 114)

PGFp02 (1st loop)
 CGGATCAATGGCGATGTTGAGATGAGTCACTCCCTGGCTGGTGTGGCTGATATGCGT
 and
 (SEQ ID NO: 115)

GGAAACACCGGCCAAGGGGTGACTGTCAATCCAAGTACAGATGCTGGATTGATGCCAACAGGGTT
 (SEQ ID NO: 116)

-continued

FGP₂ (2nd loop)
 AAGCTGTATCTCACATGKATCGATGAGTACTCCCTGTTTGCCTGACATCACCG (SEQ ID NO:117)
 and

ACCCAAAAACAAAGCTGTGAGAATCGATCCATGTGCGAGATAACAGCTTTGATGACT (SEQ ID NO:118)

FGP₉B (2nd loop)
 AAGCTGTATCTGACATGATCGATTGGACACCTTGTGTTTGCCTGACATCACCG (SEQ ID NO:119)
 and

ACGCCAAAAACAAAGCTGTGCAATCGATCCATGTACAGATAACAGCTTTGATGACT (SEQ ID NO:120)

189 (2nd loop)
 CAAAAGCTGCGCATGTGTTACTACAGATGGATCGAGTGTGTTTGCCTGACATCACCG (SEQ ID NO:121)
 and

ACGCCAAAAACAAAGCTGTGCAATCGATCCATGTACAGATAACAGCTTTGATGACTATG (SEQ ID NO:122)

1A12 (2nd loop)
 CAAAAGCTGTGCTGCCAACACTTGGATCATAATGTTGTTTGCCTGACATCACCG (SEQ ID NO:123)

and

ACGCCAAAAACACATATGAGTCGAAAGTGTGGCAGGACACGCTTTGATGACTATGAC (SEQ ID NO:124)

1E11 (2nd loop)
 CAAAAGCTGCGCATGTGTTACTCTCAATTCGACAAATGTTTGCCTGACATCACCG (SEQ ID NO:125)

and

ACGCCAAAAACATTTGGGGAATTGAGAGTAAACATGCGACGTTTGATGACTATG (SEQ ID NO:126)

TGP₉1 (2nd loop)
 CAAAAGCTGTGCTGGGGAAAGCTGATCGTTCTCTGTGTTTGCCTGACATCACCG (SEQ ID NO:127)

and

TGTCGACGCAATTACAGGNGGAAAGCTTATCGTTTCCGGACAAAGCAUCTTTGATCUCACTATUAC (SEQ ID NO:128)

The DNA sequences of the oligonucleotide pair used to make the cassette to introduce the MM021 peptide into the chymotrypsin reactive site loops of the pJ2M103-lnk2-BBI³⁵ expression vector are provided below. The cassette was ligated into the SphI and Sall restriction sites in the vector.

MM021 (2nd loop)
 (SEQ ID NO:129)
 CAAAAGCTGTGCTGTGAAACACAAACGTCAGCTTTGTTGCG
 and
 (SEQ ID NO:130)
 TCGAGCAGAAAAACATAAAAGACGCTACGTTGTTGATCAAGCACAGC

TTTGCGATG

Libraries made of cysteine constrained peptides are popular reagents (e.g., the commercially available PhD-C7C Phage Display Peptide Library Kit; NEB) for selecting peptides that bind to substrates of interest. BBI has two cysteine constrained reactive site loops that are structurally similar to the peptide loops displayed in various methods used to select peptide binders. So, once a cysteine constrained binding peptide has been selected, BBI is suitable for use as a scaffold to present the peptide in a binding reaction.

The VegF binding peptide CK37281 (See e.g., co-pending U.S. Provisional Patent Application Ser. No. 60/520,403, filed Nov. 13, 2003, incorporated herein by reference) was grafted into BBI by replacing the trypsin, chymotrypsin, or both reactive site loops, with the CK37281 peptide sequence (ACYNLYGWTC) (SEQ ID NO:9) by using DNA oligonucleotide cassettes. To facilitate the construction, an EcoRI site was introduced in the coding region of the BBI gene (custom synthesized by Operon Technologies; See, Example 1) between the trypsin and chymotrypsin reactive site loops

by QuikChange® site-directed mutagenesis, using methods described by the manufacturer (Stratagene) using the primers BowBeco-F and BowBeco-R, shown above (0.5 pmol of each primer was used in the QuikChange® reaction; after an initial denaturation step of 97° C. for 3 minutes, 18 PCR cycles of 68° C. for 12 minutes, 95° C. for 30 seconds and 55° C. for one minute, followed by a final extension reaction for 15 minutes at 68° C.).

To replace the trypsin inhibitory peptide loop, two DNA oligonucleotides (IBBCK81+ and IBBCK81-) were annealed and ligated into the SacI and EcoRI restriction sites. Likewise, to replace the chymotrypsin inhibitory peptide loop, EcoRI and Sall sites were used for insertion of a DNA cassette made by annealing the oligonucleotides (2BBck81+ and 2BBck81-). The CK37281 peptide was grafted into both loops by inserting the CK37281 peptide in the chymotrypsin loop (using the oligonucleotides (2BBck81+ and 2BBck81-) after the trypsin loop was first replaced by the CK37281 peptide. BBI with the CK37281 peptide in the trypsin loop (IBBCK81) was moved into the pJ2M103BBI expression vector as a SacI-SphI fragment. BBI with the CK37281 in the chymotrypsin loop (2BBck81), or both loops (2BBck81), was moved into pJ2M103BBI as SacI-Sall fragments. The correct sequences were verified by DNA sequencing (the sequence of 12BBck81 gene is shown in FIG. 3). The resulting vectors, pJM103-1BBck81, pJM103-2BBck81, or pJM103-12BBck81, were used to transform *B. subtilis* BG393-4comk, and the production of the BCE fusion proteins was determined as in Example 1 above.

The fusion protein running at ~44 kDa was detected by SDS-PAGE to be the major protein present in the cell free broth. Although in some cases, there was significant degradation (up to 50%) of the BBI moiety (especially after >48 h of growth in MBD medium), as observed by the presence of

a prominent protein band running at ~34 kDa corresponding to the BCE103 catalytic core. In these cases, the titers of the BCE103 cellulase were similar to that measured with fusions to the wild-type BBI (Example 1), but the activity of the BBI (trypsin inhibition with 2BBIck81, or chymotrypsin inhibition with 1BBIck81) was generally about two fold less.

To reduce the proteolytic degradation of BBI variants during growth (i.e. decrease the amount of BCE103 cellulase core present on SDS-PAGE gels in comparison to the fusion protein), a *Bacillus subtilis* strain with nine protease genes deleted (BG6006 (degU^{0.5}32, oppA, AspP¹E3501, AaprE, AaprF, AaprG, AisP¹, AaprA, AaprY^{0.5}Yf, AaprB, amyE^{0.5}xylylPxylAcornK^{erm})), was used as an expression host, and the growth temperature (35°C) and aeration (200 rpm) were reduced. With these changes, a major fusion protein band (~44 kDa) was observed on SDS-PAGE gels with an insignificant band present at the molecular weight expected for the BCE catalytic core protein (~34 kDa).

In addition to the CK37281 peptide, a number of other cysteine constrained peptides were produced when substituted into the trypsin or chymotrypsin reactive site loops of BBI fused to the C-terminus of the BCE103 cellulase. Specific examples included:

- (1) Peptides designed or selected as complement antagonists, compstatin introduced into the 1st or 2nd reactive site loops (See, Sahu et al., *J. Immunol.* 157: 884-891, [1996]), C2c (1st loop), C3c (1st loop), C4c (1st loop) and C5c (1st loop); peptides selected in a Factor B binding reaction 1B, 2B, 4A, 5A, 6-1A, 7A, 8B, 9A, 10B, 11-1A, 12B, 13A, and 15-1A (all in 2nd loop);
- (2) Peptides designed to bind to the proteases Factor Xa or stratum corneum chymotrypsin, Xai (2nd loop) or hSCC1 (1st loop), respectively;
- (3) Peptides selected in FG5F5 binding reactions 1A6 (1st or 2nd loop), 1C2 (1st or 2nd loop), (1st or 2nd loop), 2E5 (1st, 2nd or both loops), FG1F5s (1st or 2nd loop), FG1Fkr (1st or 2nd loop), FG1H1 (1st or 2nd loop), FG1Fg (1st or 2nd loop), MM005 (1st or 2nd loop), MM007 (1st, 2nd or both loops), MM009 (2nd loop), MM010 (1st, 2nd or both loops), MM017 (2nd loop), FGFPs1 (2nd loop), FGFPs2 (1st, 2nd or both loops), and FGFPs3 (2nd loop); and
- (4) Peptides selected in TGFB-1 binding reactions 1A8 (2nd loop), 1A12 (2nd loop), 1E11 (2nd loop), TGFPs1 (2nd loop), and MM021 (2nd loop).

The oligonucleotides used to introduce these peptides into either the trypsin (1st loop) or chymotrypsin (2nd loop) reactive site loops, and methods used to graft these peptides into BBI, are provided above. In all cases, fusion proteins were produced as determined by SDS-PAGE gels. However, with some substituted peptides, the amount of intact fusion protein was increased by reducing the proteolytic degradation as described above for the CK37281 substituted peptide.

Example 3

Activation of BBI by Thiol Reducing/Oxidizing Agents

After growth, the activity of the BBI (by trypsin or chymotrypsin inhibition) is typically some 5-20 times lower than what would be expected from the activity of the BCE103 cellulase measured in the cell free supernatants (the two molecules should be present at a 1:1 molar ratio in the fusion protein). An increase in the activity of BBI (measured by either trypsin or chymotrypsin inhibition) in the BCE103-BBI fusion protein can be routinely obtained by adding bME; typically concentrations of 1-4 mM added to the MBD growth medium about 14 h after inoculation. The trypsin or chymotrypsin inhibitory activity of BBI in the fusion protein is also lower than expected when binding peptides (e.g. VegF binding peptide CK37281) replace the chymotrypsin or trypsin reactive site loop, respectively. As with the wild-type BBI, the inhibitory activity can be increased by treatment with bME. Unexpectedly, other thiol reducing agents (e.g., cysteine, reduced glutathione, DL-dithiothreitol and Tris[2-carboxyethyl]phosphine) had small or negligible effects on the activation of BBI during growth in these experiments. Also, additions of antioxidants (e.g., ascorbic acid or DL-*α*-tocopherol acetate) or other adjuvants to the growth medium (e.g., isoleucine, soybean oil, Tween-80), or growth at 30°C, did not significantly improve the BCE103: BBI activity ratio.

Specifically, to determine the BBI activation during growth, cultures of *B. subtilis* BG6006 transformed with p2JMB103-E3-2BBIck81 (See, Example 4, below) were grown in 40 mL MBD medium in 250 mL shake flasks at 37°C for 13 h. Then, the thiol reducing agents indicated on the graph in FIG. 4 were added and cell sup emulsants harvested after 62 h of growth. The reagents 2-mercaptoethanol (BME), cysteine (Cys), reduced glutathione (Glut), and DL-dithiothreitol (DTT) were added to the growth medium to the final concentrations indicated on the graph provided in FIG. 4. Concentrations of 5 mM BME can result in better BCE 103: BBI activity ratios, but typically result in an overall decrease in both BCE103 and BBI titers (see FIG. 4), at least partially due to the reduction in bacterial growth caused by the added reagent. Titers of BCE103 and 2BBIck81 were determined using the assays described in Example 1.

BBI activation was also achieved after partial purification of the fusion proteins (e.g. BCE-Ink2-2BBIck81, see Example 4 below) by Q-Sepharose ion exchange chromatography.

The fusion protein was purified from cell free broth obtained from shake flasks or fermentor runs. The broth was filtered, diluted five to ten fold in water, and the pH adjusted to pH 7.5-8.0. The diluted sample was loaded onto a column packed with Q-Sepharose resin (GE Healthcare). The column was washed with 50 mM Tris pH 7.5 and then washed again in the same buffer containing 300 mM NaCl. The fusion protein was eluted in the same buffer with 700 mM NaCl.

To activate the BBI, the pooled fusion protein fractions were diluted ten fold in Assay Buffer then treated with 2 mM β ME and 0.2 mM oxidized glutathione (GSSG) with constant mixing on a stir plate or rocker platform for about 24 h at room temperature. The BBI could generally be activated to about 70-100% of the expected trypsin inhibitory activity based on the measured concentration of the BCE103 cellulase. Although the activation method outlined above generally yielded the best results, in some cases, in order to maximize the activation of a given sample, screens were performed in 96-well plates to determine the optimal conditions. Initially, the typical conditions screened were the dilution in Assay Buffer (e.g., a 2-50 fold dilution series), β ME concentration (e.g., series between 0.5-5 mM) and oxidized glutathione concentration (e.g. 0 mM then a series of 1/20 to 1/2 the β ME concentration).

The activation of the fusion protein BCE-Ink2-2BBIck81 is shown in FIG. 5. In this specific example, the fusion protein from a Q-Sepharose purification was diluted 1:10 in Dulbecco's PBS (Mediatech) with 0.005% TWEEN8-80. Beta-mercaptoethanol was added to a final concentration of 3 mM and incubated overnight at room temperature on a rocker. The sample was further incubated at room temperature for about 6 h with vigorous stirring on a magnetic stir plate. The titers of the BCE103 and 2BBIck81 (before and after β ME treatment) were determined by cellulase assays and trypsin inhibitory assays, respectively.

In some embodiments, such as for activating BBI or its variants in cell free broth from large volume fermentations, it

is desirable to reduce the dilution and β ME concentration in the activation reaction. This can be accomplished by using higher concentrations of buffer (500 mM Tris pH 8.6), or changing to zwitterionic buffers such as CHES (also CAPS, Tricine, TAPS, and other suitable zwitterionic buffers). For example, cell free broth (or fusion protein fractions purified by ion exchange chromatography) was diluted 1:1 in 375 mM CHES pH 8.6 with 0.005% TWEEH®-80 then activated with 1 mM β ME and 10 mM Na₂SO₃ and incubated with stirring at room temperature for about 24 h. BBI or its variants, as BCE103 cellulase fusion proteins, were routinely activated by this method to 70-100% of the expected value (based on BCE103 cellulase activities).

Example 4

Release of Free BBI/Variants by Cleavage of the BCE103-BBI Fusion Proteins

This Example describes experiments developed to release free BBI or its variants by cleavage of the BCE103-BBI fusion proteins.

The sequences of the DNA oligonucleotide pairs that were annealed and ligated into the BamHI and SacI sites of pJEM103-BBI to generate potential cleavage sites during culture growth between the BCE103 catalytic domain and BBI are provided below.

BCEsubBBI
(*a* subtilisin-type sensitive peptide sequence)
(SEQ ID NO:131)
GATCCAGGTGGAGCTGCTTCTAGTGACGATGAGAGCT
and
(SEQ ID NO:132)
CTCATCGTCACATAAGCAGCTCCACCTG

BCEcbdBLLB1
(a portion of the 1st CBD)
(SEQ ID NO:133)
GATCCAGGTGGACCTGACCCAACTCTCCATCTGATCTTGAGAATACCC

BCEProBBI
(*the* entire pro peptide of BBI)
(SEQ ID NO:135)
GATCCGGCGRACACTGGCTCTGCTTAAGCTGGCTGCTTATGAAATCAGA
CCATCAGCACACAAATGACGATGAGAGCT
and
(SEQ ID NO:136)
CTCATCGTCATTCGCTGCTGATGGCTGATTTCTATAGCAGCGCAAGCT
CAGGGTCACTCG

BCEshortBBI
(*a* C-terminal portion of the pro peptide of BBI)
(SEQ ID NO:137)
GATCCAAAATCAGACCATCAGCACAGCAATGACGATGAGAGCT
and
(SEQ ID NO:138)
CTCATCGTCATTCGCTGCTGATGGCTGATTTCTATAGCAGCGCAAGCT

The sequences of the DNA oligonucleotide pair that was annealed and ligated into the BamHI and SacI sites of pJEM103-BBI to fuse BBI to the 2nd CBD linker of BCE103 cellulase are provided below.

BCEcbdIDBBI
(SEQ ID NO:139)
GATCCAGGAGAGCGCAACCGCCCCAAAGTGATCAGGGAGGTATCC
10 ACATGGATTCAAAATCAAAATTACACAAATGAAATTGTATCATACCG
GTCAGTATGGCAAGCAGAAATGGTGACACAAAAATCAGAGGCCAGGTGAG
CCATACGGTCTGGAAACACTCAAAATCTGACCCAGATTCAGAGGA

15 **TGAGAGCT**
and
(SEQ ID NO:140)
CTCATCGTCGAATCTGGTCAGATTGAGTTGGTCCACGAGGACCTATG

20 GGTCACTCGCTCTGATTTGTCACCTTGTGAAATTGATTGATTCATCCATGC
CCGTTATGATACACRATTTCATTGTGAAATTGATTGATTCATCCATGC
TGTTGACTCTCTGGATCACTTGGGGCGTGGGTCCGGTCTCTCTG

25 The peptide sequences susceptible to acid cleavage between aspartic acid and proline residues are provided below.

30 **Linker 1-** (SEQ ID NO:141) (Uidel et al., *J. Biol. Chem.* 278:13944-51 [2003])
Linker 2- (SEQ ID NO:142) (Seglas et al., *FEBS Lett.* 371:171-175 [1995])
35 **Linker 3-** (SEQ ID NO:143) (Kemperman et al., *Appl. Env. Microbiol.*, 69: 1589-1597 [2003])

40 Oligonucleotide primers used to introduce a BssHII site into pJM 103(BBI) by QuikChange® site-directed mutagenesis are provided below.

45 **BCEbbs-P 5'-** TGGCGTCTGGCACATGAGGGCGCGCTGATGATTA
(SEQ ID NO:144)
BCEbbs-R 5'- TAATCATCAGCTCTGGCGCTCATGTTGGACGCC
(SEQ ID NO:145)

50 Sequences of the DNA oligonucleotides that were annealed as a cassette (SalI-HindIII) to introduce HindIII and XbaI sites after the stop codon of BBI, to introduce a Pael site after the LAT, and remove the original 11 HindIII site are provided below.

55 **BCEter+**
(SEQ ID NO:146)
5'-GACATCACGACTTCTGCTATGACGGCATGTAACCAAGCGAGCA

TAAGAGCTTAAGCTTAACCTCGAGGTTAACAGGAGCGGATTTCCTG
60 AAGGAATCCGTTTTTATTTTAATTAG

65 **BCEter+**
(SEQ ID NO:147)
5'-AGCTCTTAATAAAAATAAAACCGGATTCTTCAGGAATCCGTC
CTCTGTTAACCTCGAGGTTAACGTTTTAGTTCTCTTATCGTCCTGGCTG
GTTACATCGCTCATAGCAGAGTCCTGGATG

PCR primers used to generate the acid labile linkers provided above (i.e., Linker 1, Linker 2, and Linker 3) inserted between the BCE103 catalytic domain and BBI are provided below.

BCE103coreBeeHII_FW
5'-CAGCAACATGAGCGCAGGCTG (SEQ ID NO:148)

LinkerWDPHI_RV
5'-ATCGCTGATCCGGATAGTGGGATCTCCCAAGATGCTGATTC (SEQ ID NO:149)

TTATTTTTCCCC

LinkerDNDPI_FW
5'-ATCGCTGATCCGGATAGGGATCATGGTGTAGATGCTGATTC (SEQ ID NO:150)

TTATTTTTCCCC

LinkerVVADPN_FW
5'-ATCGCTGATCCGGATGGGATCTGAACTACAGATGCTGATTC (SEQ ID NO:151)

TTATTTTTCCCC

PCR primers used to generate the acid labile linkers provided above (i.e., Linker 1, Linker 2, and Linker 3) inserted into the 1st CBD linker.

BCE103corePstI_FW
GCATAAGGT GAGTCATCT CGCGC (SEQ ID NO:152)

LplusWDPHY_FW
5'-ATCGCTGATCCGGATAGTGGGGCTCTCCACGGTCTCTGGAT (SEQ ID NO:153)

CAGATGGCGG

LplusDNDPI_FW
5'-ATCGCTGATCCGGATGGGATCATGGTGTGGGATCTCTGGAT (SEQ ID NO:154)

CAGATGGCGG

LplusVVADPN_FW
5'-ATCGCTGATCCGGATGGGATCTGAACTACCGGATCTCTGGAT (SEQ ID NO:155)

CAGATGGCGG

Protein sequence of the acid labile linkers inserted between the BCE103 catalytic domain and BBI are provided below. The acid labile linkers are shown in bold type and the sequences from the first CBD domain are underlined.

Linker 1
BCE-WD^{DPHY}-PPD-BBI (SEQ ID NO: 156)

Linker 2
BCE-DND^{PI}-PPD-BBI (SEQ ID NO: 157)

Linker 3
BCE-VVAD^{DPN}-PPD-BBI (SEQ ID NO: 158)

LinkerPlus 1
BCE-IPPSDTPPSDGP-WD^{DPHY}-PPD-BBI (SEQ ID NO: 159)

LinkerPlus 2
BCE-IPPSDTPPSDGP-DND^{PI}-PPD-BBI (SEQ ID NO: 160)

LinkerPlus 3
BCE-IPPSDTPPSDGP-VVAD^{DPN}-PPD-BBI (SEQ ID NO: 161)

The sequences of the DNA oligonucleotide pairs that were annealed and ligated into the BamHI and SacI sites of pJM103-BBI to generate potential cleavage sites between the BCE103 catalytic domain and BBI during the purification process are provided below.

BCEentBBI
(Enteropeptidase cleavage site)
5' GATCCAGGTGGAGACGACATGACAAGACGATGAGAGCT and (SEQ ID NO:162)

CTCATCGCTTGTCTCGTCCTCACCTG (SEQ ID NO:163)

BCEgenenBBI
(Genenase I cleavage site)
5' GATCCAGGTGGCTCTCATIACGACGATGAGAGCT and (SEQ ID NO:164)

CTCATCGCTTAATGAGCACGACCTG (SEQ ID NO:165)

The sequences of the DNA oligonucleotide pairs that were annealed and ligated into the BamHI and SacI sites of pJM103-Ink2-2BBIck81 to generate potential cleavage sites between the BCE103 catalytic domain and BBI during the purification process are provided below.

BCEfurInBBI
(Furin/Blisterase cleavage site)
5' GATCCAGGTGCTAAAAAGACGATGAGAGCT and (SEQ ID NO:166)

CTCATCGCTTCTTGTAGCACGTG (SEQ ID NO:167)

BCEgenenBBI
(Genenase I cleavage site)
5' GATCCAGGCCTGCACACTACAACGACGATGAGAGCT and (SEQ ID NO:168)

CTCATCGCTTGTAGTGTGAGCGCCTG (SEQ ID NO:169)

BCEfleBBI
(Mpr cleavage site)
5' GATCCATTCTTGTAGACGATGAGAGCT and (SEQ ID NO:170)

CTCATCGCTTCAAGGAATG (SEQ ID NO:171)

Sequences of the oligonucleotide primer pairs used to introduce the E and E3 linkers in Linker 2 by QuikChange site-directed mutagenesis (Stratagene) are provided below.

BCE-Elnk-BBI
(Mpr cleavage site)
5' CCCATACCGGAAGCAGACGATGAGAGCT and (SEQ ID NO:172)

CATCGCTGGCTCCGGATGGGATCATGGT (SEQ ID NO:173)

The protein sequence of the E3 linker between the BCE103 catalytic domain and BBI was DNNIDIPPEPDDESFM-MPIEP (SEQ ID NO:174). In this sequence, the E Linker is underlined and the sequence generated by faulty recombination in *E. coli* is shown in bold type. Cleavage by Mpr (or V8 protease) can occur after any of the three glutamic acids present in the E3 Linker. Thus, the structure was BCE-(SFQ ID NO:174)-BBI.

The sequences of the DNA oligonucleotide pairs that were annealed and ligated into the BamHI and SacI sites of pJM103-Ink2-2BBIck81 to generate potential Genenase I cleavage sites between the BCE103 catalytic domain and BBI are provided below.

BCEgenase3BBI
 (SEQ ID NO:175)
 GATCCAGGCGCTGCACACTACAATCAGACCATCAGCACAGCAATGACGA
 and
 (SEQ ID NO:176)
 CTCATCGTCATTGCTGTGCTGTGATTTGTTAGTGTGTCAGCCCTG
 BCEgenase4BBI
 (SEQ ID NO:177)
 GATCCAGGCGCTGCACACTACGTGATTTGTTAGCAGATGAGACT
 and
 (SEQ ID NO:178)
 CTCATCGTCATTGAAATTCTACGTAGTGTGTCAGCCCTG

The protein sequence of a Genenase I sensitive cleavage site (also acid and Mpr sensitive) inserted between the BCE103 catalytic domain and BBI was DNNNDPIPDP-GAAHYVEFQ (SEQ ID NO:179). The Genenase I site (Gen4 Linker) is bold type (cleavage occurs between the tyrosine and valine) (NEB) and Linker 2 is underlined. Cleavage by Mpr can also occur after the glutamic acid that follows the valine in the Gen4 linker. The sequence used herein was BCE-SFQ ID NO:179(BBI).

Cleavage sites in the BCE103-link2-2BBIck81 fusion protein are indicated below. The C-terminal seven amino acids of the BCE103 catalytic domain (underlined), linker 2 sequence (bold type), and 2BBIck81 sequences are shown. The acid/heat labile Asp-Pro bonds are indicated with solid headed arrows and the Mpr sensitive bonds after glutamic acids are indicated with line headed arrows.

▼ (SEQ ID NO:180)
 VTKREASDNDNDPIPDPDESSVPCCDQCACTKSNPQCRCSMDML
 ▼
 HSCHSACKSCACSYNLIGWTCFCVDDTDPCTYEPCKPSEDDKEN

In order to isolate free BBI or its variants, the BBI moiety needs to be cleaved from the BCE103-BBI fusion protein. In some embodiments, this is accomplished during growth, by proteases intrinsically produced by *B. subtilis*. In some alternative embodiments, this cleavage occurs after growth, during the purification process (e.g., by acid/heat or proteolytic cleavage). Linkers potentially susceptible to cleavage during growth were designed (See, above, sub, cbdB, pro, shortpro, and cbdD) and cloned into the pJM103BBI or p2JM103BBI expression vectors as BamHI-SacI cassettes. The production of fusion protein versus BCE103 catalytic domain was analyzed on SDS-PAGE gels as described in Example 1.

Little cleavage of the fusion protein was observed for all these linkers except with the pro linker, which was nearly completely cleaved so that very little intact fusion protein was observed on gels, although there was a large band corresponding to the BCE103 catalytic core. Unfortunately, this cleavage during growth resulted in negligible BBI activity measured in cell free supernatants and no BBI band could be identified on SDS-PAGE gels. Although it is not intended that the present invention be limited to any particular mechanism or theory, it is possible that the BBI is particularly sensitive to proteolytic degradation in its inactive form. Thus, cleavage during the purification process after activation is generally preferred.

In some embodiments, the bonds between aspartic acid and proline residues are cleaved by heat treatment at acidic pH as known in the art (See e.g., Landon. Meth. Enzymol., 47:145-

149 [1977]). The L¹ CBD linker in the BCE103 cellulase has three Asp-Pro dipeptide sequences (See, FIG. 1) with the potential to be cleaved by acid/heat treatment. However, cleavage by acid/heat treatment at these sites was found to be inefficient. Protein sequences that are especially labile to acid/heat have been described in the literature, three of such sequences are WGDPHY (SEQ ID NO:141), DNNNDP (SEQ ID NO:142), and VVADPN (SEQ ID NO:143)(i.e., Linkers 1, 2 and 3).

Before these acid labile linkers were introduced into the BCE103-BBI expression vector, pJM103BBI, a BssHIII site was introduced by QuikChange® XL (Stratagene) mutagenesis (using the manufacturer's methods; and described in Example 2 above, except 8 minute extension and 1 minute denaturation steps were used) in the aprE signal sequence coding region using the oligonucleotide primers BCEBbs-F and BCEBbs-R (provided above). Then, HindIII and XbaI sites were inserted in front of the LAT terminator (after the BBI stop codon) and a Pael site was added after the terminator (the original HindIII site after the LAT terminator was removed) by inserting an oligonucleotide cassette (BCEerm4+ and BCEerm1-, provided above) into the Sall and the original HindIII sites. This new vector was called "p2JM103BBI".

The acid labile linker fragments were generated by PCR, using forward primer BCE103coreBssHIII_FW with each of the reverse primers, linker WGDPHY_R, linker DNNNDP_R, or linker VVADPN_RV (the sequences of which are all provided above) and p2JM103BBI as the template (see Example 1 for the PCR protocol). The PCR fragments of 970 bp were digested with BamHI and PstI, the 154 bp fragments encoding the acid linker fragments were isolated from an agarose gel after electrophoresis, and ligated into the p2JM103 vector digested with BamHI and PstI that had also been purified from a gel. The linker sequences in the final expression vectors, p2JM103lnk1-BBI, p2JM103lnk2-BBI and p2JM103lnk3-BBI, were verified by DNA sequencing.

Competent *B. subtilis* strain BG3934comK or BG6006 were transformed with the plasmids, colonies selected on 5 µg/ml chloramphenicol LA plates and amplified to 25 µg/ml chloramphenicol as described in Example 1.

Similarly, the acid labile linkers were inserted into the first CBD linker. Specifically, PCR fragments were generated using the forward primer BCE103corePstI_FW with the reverse primers LplusWGDPHY_RV, LplusDNNNDP_RV, or LplusVVADPN_RV (See above, for the sequences) with p2JM103BBI as a template. The PCR fragments of about 150 bp were digested with BamHI and PstI, purified and ligated to the p2JM103BBI vector digested with BamHI and PstI. The correct sequences were verified by DNA sequencing and the plasmids p2JM103lnk1-BBI, p2JM103lnk2-BBI and p2JM103lnk3-BBI were used to transform *B. subtilis* strains as described above.

After growth in MBD medium, the fusion protein were purified by ion exchange chromatography essentially as described above (See, Example 2). The fusion protein was cleaved by treatment at 55° C. for 16 h in 10% formic acid. The BCE103 catalytic domain precipitated during the acid treatment and was removed by centrifugation. The free BBI in the supernatant was dried overnight on a SpeedVac. The sample was suspended in 50 mM Tris pH 8 before loading on the SDS-PAGE gel. By analysis of the protein loaded on SDS-PAGE gels, it was observed that acid cleavage was much more efficient in the fusion proteins where Linker 2 was inserted between the BCE103 catalytic domain and BBI (BCE-DNNNDP-PDP-BBI). This linker was found to be cleaved in a couple of hours at 75° C. in 20 mM glycine pH 2.

In alternative embodiments, the fusion protein was cleaved by treatment with a protease during the purification process. Linkers were designed with cleavage sites for glutamic acid specific proteases (e.g., Mpr or V8 protease). Furin/blisterase, Genenase I, and Enteropeptidase (Enterokinase). These linkers were introduced as oligonucleotide cassettes (See above, for the sequences) between the BCE103 catalytic core and BBI in the expression vector using the BamHI and SacI sites (See, FIG. 1). In the coding region of the original expression vector (pJM103(BBI)), there is a glutamic acid residue in the 1' CBD domain and at the third residue in BBI (See, FIG. 1), which is contemplated to be susceptible to cleavage by glutamic acid specific proteases such as *B. subtilis* Mpr (BsMpr) or V8 protease. However, neither BsMpr nor V8 protease were found to cleave the BCE-BBI fusion protein very efficiently at these sites. Thus, it was necessary to design other linkers that were susceptible to cleavage by these proteases.

The six acid labile linkers described above were tested for cleavage by BsMpr. These fusion proteins were cleaved by treatment for 16 h with 16 μ g of BsMpr at room temperature. After cleavage, the BCE103 catalytic domain was precipitated by the addition of 10% formic acid and removed by centrifugation. The free BBI in the supernatant was dried overnight on a SpeedVac. The sample was suspended in 50 mM Tris pH 8.8 before loading on the SDS-PAGE. Similar to the acid cleavage, the BCE-DNNPDI-PDP-BBI (Linker 2) fusion protein was much more efficiently cleaved by BsMpr than any of the other linkers. Therefore, BBI and its variants were found to be effectively released from the BCE-DNNPDI-PDP-BBI fusion protein either by acid/heat treatment or proteolytic digestion with a glutamic acid specific protease such as Mpr. Several other linkers designed for cleavage by Mpr (e.g., F, F3 linker, and fl, provided above) were tested but none of them had any advantages over Linker 2 (the E3 linker was generated by Eutly recombinination in *E. coli* after transformation with the QuikChange® site-directed mutagenesis reaction designed to construct the E linker). As shown above, there are two acid/heat labile cleavage sites in Linker 2 and three sites sensitive to cleavage by Mpr.

Linkers designed for cleavage by Furin or Blisterase (NEB) (BCE:furin(BBI)), or Enteropeptidase (Enterokinase, NEB) (BCE:en(BBI)) were tested, but none of these sequences were cleaved efficiently by the appropriate protease. Four linkers were also designed (BCE:genen(BBI), BCE:genen2(BBI), BCE:genen3(BBI), and BCE:genen4(BBI)) and tested for cleavage by Genenase I (NEB). Efficient cleavage of the fusion protein was observed only with the Gen4 Linker (BCE:genen4(BBI)). BsMpr was also found to efficiently cleave the Gen4 linker.

After activation of the purified BCE:lnk2-2BBIck81 fusion protein, cleavage by BsMpr does not go to completion as judged by SDS-PAGE gels. However, it was discovered that complete cleavage after activation of BCE-BBI fusion proteins with Linker 2 (or the Gen4 linker) can be accomplished by using the Mpr protease isolated from *Bacillus licheniformis* (BIMpr). While it is not intended that the present invention be limited to any particular mechanism, cleavage after the third amino acid in mature BBI appeared to be more sensitive to BIMpr while cleavage after the sixth amino acid from the C-terminus of BBI is more sensitive to BsMpr cleavage.

In some embodiments, after cleavage, the BBI is purified away from the BCE103 catalytic domain by selective acid precipitation (pH 3 or lower) of the BCE103 catalytic domain as described above, ion exchange chromatography (See, Example 5), or by selective binding of BBI on an anhydrot-

rypsin-agarose (Sigma) column loaded in 50 mM Tris pH 8.0, washed with 50 mM Tris pH 8.0 with 150 mM NaCl, then eluting bound BBI with 50 mM glycine pH 2.2 with 300 mM NaCl).

Example 5

Binding of BBiick81 to VegF

In this Example, experiments conducted to assess the binding of BBiick81 to VegF are described. The BCE103-linker-2BBiick81 fusion protein was produced in *B. subtilis* as described in Example 2. The fusion protein was purified, and the BBI trypsin inhibitory activity was increased by treatment with BME and oxidized glutathione as described in Example 3. The fusion protein was cleaved by BsMpr protease (See, Example 4) and the free 2BBiick81 was purified from the BCE103 catalytic domain by ion exchange chromatography using a Q-Sepharose column.

Briefly, after cleavage, the pH of the cleaved sample was adjusted to 5.5, the sample was then loaded onto the column (equilibrated with 25 mM MES pH 5.5). The free 2BBiick81 was washed through the column using 25 mM sodium acetate pH 5.0 while the BCE103 catalytic core remained bound to the resin. The 2BBiick81 fraction was concentrated by ultrafiltration and analyzed using an electrochemiluminescence (ECL) based binding assay (BioVeris). The Anti-VegF antibody (Santa Cruz) and VegF (PeproTech) were labeled with the electrochemiluminescent dye and biotin, respectively, as described by the manufacturer (BioVeris). All materials were in Dulbecco's PBS (Mediatech) supplemented with 0.1% Tween®-80. An initial dilution series of Anti-VegF antibody (125, 250 and 500 ng/ml) and VegF (100, 150, 200 and 250 ng/ml) were tested in the binding assay to determine the concentrations of each that would give a robust ECL signal.

For testing 2BBiick81 binding, 50 μ l of 500 ng/ml ECL labeled Anti-VegF antibody, 50 μ l of 250 ng/ml biotinylated VegF and 100 μ L 2BBiick81 (series of 12.5, 15, 31.25, 62.5, 125, 250 or 500 ng/ml) were incubated at room temperature for 2 h with shaking. Then, 50 μ l of 0.2 mg/ml streavidin coated beads were added and the reaction was incubated at room temperature for 30 minutes. The ECL signal was measured using a BioVeris M8/384 Analyzer as described by the manufacturer (BioVeris). As shown in FIG. 6, the ECL signal decreased as increasing concentrations of 2BBiick81 displaced more of the labeled Anti-VegF antibody bound to VegF attached to the magnetic beads.

Thus, the CK37281 peptide when grafted onto the chymotrypsin inhibitory loop of BBI (2BBiick81) competed with the Anti-VegF antibody for binding to VegF at micromolar concentrations. In fact, 2BBiick81 competed for VegF binding better than the synthesized CK37281 peptide itself (See, FIG. 6). The CK37281 peptide inserted into the trypsin inhibitory loop, BBiick81, also competed with the Anti-VegF antibody in the BioVeris assay. Thus, BBI was found to be useful as a scaffold to present active binding peptides selected by various screening methods.

Example 6

Use of Alternative Fusion Partners for the Production of 2BBiick81

In this Example, experiments conducted to evaluate alternative fusion partners are described. The DNA sequence of the oligonucleotide primers used to amplify the dsbC gene (*E. coli*) from pET-40b(+) are provided below. These primers

generate a BssHII site at the 5' end and a BamHI at the 3' end for cloning into p2JM103-Gen4-2BBlck81.

DsbCBBI-F (SEQ ID NO: 181)
AACATGAGCGCGCAGGGCTGATGACGGGGCAATTCAACAAAGCTTAA
DsbCBBI-R (SEQ ID NO: 182)
TGTCTCGATCCGGTATGGATCATTGGTCACCGACRACACTAGTTGA
TCTTTTACCGCTGGCATTTTTGGTG

The DNA sequences of the oligonucleotides that were annealed together to make a cassette (Alw441-BamHI) for fusing the *P. mendocina* cutinase gene to BBI with Linker 2, are provided below.

CutinaseBBI+ (SEQ ID NO: 183)
TGCACCTTCTGCTTGGCTGTGACAGAGGCTTGACAAACATGA
TCTTATTCGG
CutinaseBBI (SEQ ID NO: 184)
GATCGGGAAATAGGATCATTTGTCAAGACCTCTGCGTCACRGCACCA
ACAGAGAGAG

Because the BBI moiety has seven disulfide bonds, it is contemplated that higher titers of active BBI will be obtained using fusion proteins other than the BCE103 cellulase catalytic domain. For example, in some embodiments, compositions such as thiol-disulfide oxidoreductases and/or protein disulfide isomerases find use as fusion proteins to help produce correctly folded BBI moieties. In this embodiment, no additional activation step is needed under most circumstances. In additional embodiments, other proteins produced at high titers in *B. subtilis* also find use as fusion partners. For example, the thermostable protein disulfide isomerase from the fungus *Humicola insolens* (hPDI) has been used as a fusion partner to produce the light chain of immunoglobulin G (2 disulfides) in *Bacillus brevis* (See, Kajino et al., *Appl. Env. Microbiol.*, 66:638-642 [2000]).

To determine whether hPDI could be a better fusion partner than BCE103 for the production of BBI, this hPDI gene was synthesized (DNA2.0) and cloned into the expression vector, p2JM103-Ink2-2BBlck81 (See, Example 4) as a BssHII-SacI fragment. In designing the synthetic gene, codons occurring with high frequency in highly expressed *B. subtilis* genes were selected except in cases where restriction sites were introduced or deleted. In the final construction, the N-terminus of the mature hPDI gene was fused to the AprE signal sequence and the C-terminus was fused to a linker with an Enteropeptidase cleavage site (Kajino et al., *Appl. Env. Microbiol.*, 66:638-642 [2000]), which in turn was fused to 2BBlck81 (See, FIG. 7). This expression vector, p2JM-PDI-FK-2BBlck81, was used to transform *B. subtilis* BG6006 and the production of the fusion protein was determined in MBD medium (as described in Example 1) with or without 2 mM β ME added 14 h after inoculation.

As determined by SDS-PAGE gels, the production of the PDI-2BBlck81 fusion protein was typically somewhat less than the BCE-2BBlck81 grown under identical conditions. The BBI titers (trypsin inhibition) measured from the PDI-2BBlck81 cell free supernatants were also typically less than the BCE-2BBlck81 fusion. As with fusions to BCE103, the measured activities of BBI when fused to PDI were higher when grown in 2 mM β ME and the BBI activity was increased by the addition of β ME to the cell free supernatants after growth when grown in β ME free medium (as described in Example 3). Thus, the thiol-disulfide oxidoreductase activity

of PDI does not seem to significantly improve the titers of active 2BBlck81 in the fusion protein or obviate the need for activation of the BBI molecule.

In order to increase the reduction potential of the fusion protein, which was contemplated to improve the BBI titers during growth, DsbC from *Escherichia coli* was used as a fusion partner for 2BBlck81. The dsbC gene was amplified by PCR using Hercules Enhanced DNA polymerase as described by the manufacturer (Stratagene) using DsbC-F and DsbC-BBI-R as primers (sequences shown above) and pET-40b(+) (Novagen) as a template. The isolated PCR fragment was cloned into the vector p2JM103-Gen4-2BBlck81 (See, Example 4) as a BssHII-BamHI fragment. The correct sequence of the fusion gene was verified by DNA sequencing. In this case, the titers of the DsbC-2BBlck81 fusion protein were significantly lower than the BCE-2BBlck81 fusion protein as judged on SDS-PAGE gels and the titers of the active 2BBlck81 measured by trypsin inhibition were much lower as well.

Other proteins that are produced at high titers in *B. subtilis* find use as fusion partners for the production of BBI. One such protein is the cutinase from *Pseudomonas mendocina*, which has been expressed at high titers utilizing the aprE promoter from *B. subtilis* (See e.g., U.S. Pat. No. 5,429,950, herein incorporated by reference). The aprE-cutinase gene fusion as an EcoRI-Alw441 fragment (from pAK-15) was ligated with an Alw441-BamHI linker oligonucleotide cassette (See, sequence above) into the p2JM103-Ink2-2BBlck81 (See, Example 4) that had been cut with EcoRI and BamHI. This cutinase-linker2-2BBlck81 expression vector (See, FIG. 8 for the EcoRI-BamHI aprE-cutinase-linker2 sequence) was used to transform *B. subtilis* BG6006 cells and the fusion protein was produced in MBD medium as described previously for the other fusion proteins (See, Example 1). In this case, the cutinase-linker2-2BBlck81 fusion protein was not the major band observed on SDS-PAGE gels and the measured lipase titers (as measured using the methods provided in U.S. Pat. No. 5,429,950) and BBI titers were much less (ca. 20 fold) than found with the BCE-2BBlck81 fusion protein. Also, the BBI titers in the cutinase fusion protein were not improved significantly when 3 mM β ME was added to the growth medium. Thus, the highest titers of active 2BBlck81 was consistently obtained by activation of the BCE-2BBlck81 fusion protein. Nonetheless, it is contemplated that various fusion partners will find use in the present invention.

All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains. All patents and publications are herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

Having described the preferred embodiments of the present invention, it will appear to those ordinarily skilled in the art that various modifications may be made to the disclosed embodiments, and that such modifications are intended to be within the scope of the present invention.

Those of skill in the art readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The compositions and methods described herein are representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. It is readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

The invention illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and varia-

tion of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention as defined by the appended claims.

The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 184

<110> SEQ ID NO 1
<111> LENGTH: 1950
<112> TYPE: DNA
<113> ORGANISM: Artificial Sequence
<20> FEATURE
<223> OTHER INFORMATION: sequence encoding fusion peptide of SEQ ID NO:2

<400> SEQUENCE: 1

```

atattccat ttcttcgtc tatcaamata acagactgt gatattccaa acggagtttc 60
aaaaaagct ctgcoccttg caaatggat ggcgtctat aaaaatccg atatggta 120
aacgggggg caatggggc cgcatcggt gtcgttgtt ggcggatgt cattttat 180
cttcctccct ctatccat ttccatgtt cttccatgtt tgcggaaatgt atttttccgaa 240
atccatgttccatgtt caatccatgtt tgcggaaatgt atttttccgaa 300
caccgggggtt tccatgttccatgtt gacaaatgt tgcggggatgt cggggatgtt 360
taaaccaaaaa aacgtatgaca ttccagcata atggccatgtt actatcgatgtt atttttgttc 420
ttttttgtat gaaatgtatgtt atttccatgtt tccatgttccatgtt tgcggggatgtt 480
aaatatggat aaaaatccatgtt caaaaaaaatgtt ggcgtctactaa atattttatccatgtt 540
aaatatggatgtt caatccatgtt ttccatgttccatgtt gtcgtactgtt atttttttttaa aacgggggggg 600
gtaaatggatgtt gggggaaaatgtt aatccatgtt cggctgtgtt ttgggttata ctttgcgtt 660
tacgatgggg tccatgttccatgtt tgcgtggcgtt ggcgtgtatgtt tttccatgtt 720
tggggcaacta atttatgttccatgtt acggggatgtt atccatgttca gggggcaacta aatgttccatgtt 780
aaatggatgtt agttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata atggggatgtt 840
gtttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata atggggatgtt 900
agggggatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 960
gtatggatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1020
tatataataaa gaaatggccgtt aatggatgttccatgttccatgtt gttccatgttccatgtt 1080
tccatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1140
tcaatataataaa cccatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1200
catgttccatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1260
gttccatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1320
tttacatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1380
ggggggatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1440
tgacatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1500
gtccatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1560
atccatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1620
cgatccatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1680
ctgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1740
ggtgtgtgtatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1800
ttaatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1860
cgatccatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1920
aggaaatccatgttccatgttccatgtt gttccatgttccatgtt gtcgtggcgtt ttgggttata 1950

```

<110> SEQ ID NO 2

<111> LENGTH: 428

<112> TYPE: PRT

<113> ORGANISM: Artificial Sequence

<20> FEATURE

<223> OTHER INFORMATION: synthetic fusion peptide

<400> SEQUENCE: 2

```

Met Arg Ser Lys Leu Ile Ser Leu Leu Phe Ala Leu Thr Leu
1 5 10 15
Ile Phe Thr Met Ala Phe Ser Asn Met Ser Ala Gln Ala Asp Asp Tyr
20 25 30
Ser Val Val Glu Glu His Gly Gln Leu Ser Ile Ser Asn Gly Leu

```

-continued

35	40	45
50	55	60
Gly Leu Glu Trp Tyr Gly	Gln Phe Val Asn Tyr Glu Ser Met Lys Trp	
65	70	75
Leu Arg Asp Asp Trp Gly Ile Thr Val Phe Arg Ala Ala Met Tyr Thr		80
85	90	95
Ser Ser Gly Gly Tyr Ile Asp Asp Pro Ser Val Lys Glu Lys Val Lys		100
100	105	110
Glu Thr Val Glu Ala Ala Ile Asp Lys Glu Ile Tyr Val Ile Ile Asp		115
115	120	125
Trp His Ile Leu Ser Asp Asp Asp Pro Asn Ile Tyr Lys Glu Glu Ala		130
130	135	140
Lys Asp Phe Phe Asp Glu Met Ser Glu Leu Tyr Gly Asp Tyr Pro Asn		145
145	150	155
Val Ile Tyr Glu Ile Ala Asn Glu Pro Asn Gly Ser Asp Val Thr Trp		165
165	170	175
Asp Asn Gln Ile Lys Pro Tyr Ala Glu Glu Val Ile Pro Val Ile Arg		180
180	185	190
Asp Asn Asp Pro Asn Asn Ile Val Ile Val Gly Thr Gly Thr Trp Ser		195
195	200	205
Gln Asp Val His His Ala Asp Asn Gln Leu Ala Asp Pro Asn Val		210
210	215	220
Met Tyr Ala Phe His Phe Tyr Ala Gly Thr His Gly Glu Asn Leu Arg		230
230	235	240
Asp Glu Val Asp Tyr Ala Leu Asp Glu Gly Ala Ala Ile Phe Val Ser		245
245	250	255
Glu Trp Gly Tyr Ser Ala Ala Ala Thr Gly Asp Gly Gly Val Phe Leu Asp		260
260	265	270
Glu Ala Glu Val Trp Ile Phe Asp Met Asp Glu Arg Asn Leu Ser Trp		275
275	280	285
Ala Asn Trp Ser Leu Thr His Lys Asp Gly Glu Ser Ser Ala Leu Met		290
290	295	300
Pro Gly Ala Asn Asp Pro Thr Gly Thrp Thr Glu Ala Glu Leu Ser Pro		305
305	310	315
Ser Gly Thr Phe Val Arg Gly Lys Ile Arg Glu Ser Ser Ala Ile Pro		325
325	330	335
Pro Ser Asp Pro Thr Pro Pro Ser Asp Pro Gly Glu Pro Asp Pro Asp		340
340	345	350
Asp Glu Ser Ser Lys Pro Cys Cys Asp Glu Cys Ala Cys Thr Lys Ser		355
355	360	365
Asn Pro Pro Gly Cys Arg Cys Ser Asp Met Arg Leu Asn Ser Cys His		370
370	375	380
Ser Ala Cys Lys Ser Cys Ile Cys Ala Leu Ser Tyr Pro Ala Glu Cys		385
385	390	395
Phe Cys Val Asp Ile Tyr Asp Phe Cys Tyr Glu Pro Cys Lys Pro Ser Cys		405
405	410	415
Glu Asp Arg Lys Glu Asn His		420
420	425	

```
<210> SEQ ID NO: 3
<211> LENGTH: 223
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: sequence encoding fusion peptide of SEQ ID NO:4
```

```
<210> SEQ ID NO 4
<211> LENGTH: 73
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic fusion peptide
```

```
<400> SEQUENCE: 4
Asp Pro Asp Asp Glu Ser Ser Lys Pro Cys Cys Asp Gln Cys Ala Cys
1           5           10          15
Tyr Asn Leu Tyr Gly Trp Thr Cys Arg Cys Ser Asp Met Arg Leu Asn
20          25          30          35          40
```

-continued

Ser Cys His Ser Ala Cys Lys Ser Cys Ala Cys Tyr Asn Leu Tyr Gly
 35 40 45
 Trp Thr Cys Phe Cys Val Asp Ile Thr Asp Phe Cys Tyr Glu Pro Cys
 50 55 60
 Lys Pro Ser Glu Asp Asp Lys Glu Asn
 65 70

```
<210> SEQ ID NO 6
<211> LENGTH: 504
<212> TYPE: PROT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic fusion peptide
```

```

<400> SEQUENCE: 6

Ser Ala Gln Ala Ser Asp Val Val Gln Leu Lys Lys Asp Thr Phe Asp
      5          10          15
Asp Phe Ile Tyr Thr Asp Arg Leu Val Leu Ala Glu Phe Phe Ala Pro
      20          25          30
Trp Cys Gly His Cys Lys Ala Leu Ala Pro Glu Tyr Glu Ala Ala
      35          40          45
Thr Thr Leu Ilys Glu Ilys Asn Ilys Leu Ala Val Asp Cys Cys
      50          55          60
Glu Glu Thr Asp Leu Cys Glu Gin His Gly Val Glu Gly Tyr Pro Thr
      65          70          75          80
Leu Lys Val Phe Arg Gly Leu Asp Asn Val Ser Pro Tyr Lys Gly Glu
      85          90          95
Arg Lys Ala Ala Ala Ile Thr Ser Tyr Met Ile Lys Glu Ser Leu Pro
      100          105          110
Ala Val Ser Ile Val Thr Lys Asp Asn Leu Glu Phe Lys Lys Ala
      115          120          125
Asp Tyr Ala Val Leu Val Ala Tyr Val Asp Ala Ser Asp Lys Ala Ser
      130          135          140
Ser Glu Val Phe Thr Val Glu Val Ala Glu Lys Leu Arg Asp Asn Tyr Pro
      145          150          155          160
Phe Gly Ser Ser Asp Ala Leu Leu Ala Glu Ala Glu Gly Val Lys
      165          170          175
Ala Pro Ala Ile Val Leu Tyr Lys Asp Phe Asp Glu Gly Lys Ala Val
      180          185          190
Phe Ser Glu Ilys Phe Glu Val Glu Ala Ile Glu Lys Phe Ala Ilys Val

```

-continued

195	200	205
196	201	206
197	202	207
198	203	208
199	204	209
200	205	210
201	206	211
202	207	212
203	208	213
204	209	214
205	210	215
206	211	216
207	212	217
208	213	218
209	214	219
210	215	220
211	216	221
212	217	222
213	218	223
214	219	224
215	220	225
216	221	226
217	222	227
218	223	228
219	224	229
220	225	230
221	226	231
222	227	232
223	228	233
224	229	234
225	230	235
226	231	236
227	232	237
228	233	238
229	234	239
230	235	240
231	236	241
232	237	242
233	238	243
234	239	244
235	240	245
236	241	246
237	242	247
238	243	248
239	244	249
240	245	250
241	246	251
242	247	252
243	248	253
244	249	254
245	250	255
246	251	256
247	252	257
248	253	258
249	254	259
250	255	260
251	256	261
252	257	262
253	258	263
254	259	264
255	260	265
256	261	266
257	262	267
258	263	268
259	264	269
260	265	270
261	266	271
262	267	272
263	268	273
264	269	274
265	270	275
266	271	276
267	272	277
268	273	278
269	274	279
270	275	280
271	276	281
272	277	282
273	278	283
274	279	284
275	280	285
276	281	286
277	282	287
278	283	288
279	284	289
280	285	290
281	286	291
282	287	292
283	288	293
284	289	294
285	290	295
286	291	296
287	292	297
288	293	298
289	294	299
290	295	300
291	296	301
292	297	302
293	298	303
294	299	304
295	300	305
296	301	306
297	302	307
298	303	308
299	304	309
300	305	310
301	306	311
302	307	312
303	308	313
304	309	314
305	310	315
306	311	316
307	312	317
308	313	318
309	314	319
310	315	320
311	316	321
312	317	322
313	318	323
314	319	324
315	320	325
316	321	326
317	322	327
318	323	328
319	324	329
320	325	330
321	326	331
322	327	332
323	328	333
324	329	334
325	330	335
326	331	336
327	332	337
328	333	338
329	334	339
330	335	340
331	336	341
332	337	342
333	338	343
334	339	344
335	340	345
336	341	346
337	342	347
338	343	348
339	344	349
340	345	350
341	346	351
342	347	352
343	348	353
344	349	354
345	350	355
346	351	356
347	352	357
348	353	358
349	354	359
350	355	360
351	356	361
352	357	362
353	358	363
354	359	364
355	360	365
356	361	366
357	362	367
358	363	368
359	364	369
360	365	370
361	366	371
362	367	372
363	368	373
364	369	374
365	370	375
366	371	376
367	372	377
368	373	378
369	374	379
370	375	380
371	376	381
372	377	382
373	378	383
374	379	384
375	380	385
376	381	386
377	382	387
378	383	388
379	384	389
380	385	390
381	386	391
382	387	392
383	388	393
384	389	394
385	390	395
386	391	396
387	392	397
388	393	398
389	394	399
390	395	400
391	396	401
392	397	402
393	398	403
394	399	404
395	400	405
396	401	406
397	402	407
398	403	408
399	404	409
400	405	410
391	392	411
392	393	412
393	394	413
394	395	414
395	396	415
396	397	416
397	398	417
398	399	418
399	400	419
400	401	420
401	402	421
402	403	422
403	404	423
404	405	424
405	406	425
406	407	426
407	408	427
408	409	428
409	410	429
410	411	430
411	412	431
412	413	432
413	414	433
414	415	434
415	416	435
416	417	436
417	418	437
418	419	438
419	420	439
420	421	440
421	422	441
422	423	442
423	424	443
424	425	444
425	426	445
426	427	446
427	428	447
428	429	448
429	430	449
430	431	450
431	432	451
432	433	452
433	434	453
434	435	454
435	436	455
436	437	456
437	438	457
438	439	458
439	440	459
440	441	460
441	442	461
442	443	462
443	444	463
444	445	464
445	446	465
446	447	466
447	448	467
448	449	468
449	450	469
450	451	470
451	452	471
452	453	472
453	454	473
454	455	474
455	456	475
456	457	476
457	458	477
458	459	478
459	460	479
460	461	480
461	462	481
462	463	482
463	464	483
464	465	484
465	466	485
466	467	486
467	468	487
468	469	488
469	470	489
470	471	490
471	472	491
472	473	492
473	474	493
474	475	494
475	476	495
476	477	496
477	478	497
478	479	498
479	480	499
480	481	500

```
<210> SEQ ID NO 7
<211> LENGTH: 1495
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: sequence encoding fusion peptide of SEQ ID NO:8
```

<499> SEQUENCE: 7

<210> SEQ ID NO 8
<211> LENGTH: 295

-continued

<212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic fusion peptide

<400> SEQUENCE: 8

```

  Val Arg Ser Lys Leu Trp Ile Ser Leu Leu Phe Ala Leu Thr Leu
  1           5           10          15
  Ala Ala Ser Cys Leu Ser Val Cys Ala Thr Val Ala Ala Ala Pro Leu
  20          25          30
  Pro Asp Thr Pro Gly Ala Pro Phe Pro Ala Val Ala Asp Phe Asp Arg
  35          40          45
  Ser Gly Pro Tyr Thr Thr Ser Ser Gly Ser Gly Pro Ser Cys Arg
  50          55          60
  Ile Tyr Arg Pro Arg Asp Leu Gly Cys Gly Val Arg His Pro Val
  65          70          75          80
  Ile Leu Trp Gly Asn Gly Thr Gly Ala Gly Pro Ser Thr Tyr Ala Gly
  85          90          95
  Leu Leu Ser His Trp Ala Ser His Gly Phe Val Val Ala Ala Ala Glu
  100         105         110
  Thr Ser Asn Ala Gly Thr Gly Arg Glu Met Leu Ala Cys Leu Asp Tyr
  115         120         125
  Leu Val Arg Glu Asn Asp The Pro Tyr Gly Thr Tyr Ser Gly Lys Leu
  130         135         140
  Asn Thr Gly Arg Val Gly Thr Ser Gly His Ser Gly Gly Gly Gly
  145         150         155         160
  Ser Ile Met Ala Gly Cys Asp Thr Arg Val Arg Thr Thr Ala Pro Ile
  165         170         175
  Gin Pro Tyr Thr Leu Gly Leu Gly His Asp Ser Ala Ser Gly Arg Arg
  180         185         190
  Gin Gly Pro Met Phe Leu Met Ser Gly Gly Gly Asp Thr Ile Ala
  195         200         205
  Phe Pro Tyr Leu Asn Ala Cys Pro Val Tyr Arg Arg Ala Asn Val Pro
  210         215         220
  Val Phe Trp Gly Glu Arg Arg Tyr Val Ser His Phe Glu Pro Val Gly
  225         230         235         240
  Ser Gly Gly Ala Tyr Arg Gly Pro Ser Thr Ala Trp Phe Arg Phe Glu
  245         250         255         260
  Leu Met Asp Asp Gln Asp Ala Arg Ala Thr Phe Tyr Gly Ala Gln Cys
  260         265         270
  Ser Leu Cys Thr Ser Leu Leu Trp Ser Val Glu Arg Arg Gly Leu Asp
  275         280         285
  Asn Asn Asp Pro Ile Pro Asp
  290
  
```

<210> SEQ ID NO: 9
 <211> LENGTH: 10
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 9

```

  Ala Cys Tyr Asn Leu Tyr Trp Thr Cys
  1           5           10
  
```

<210> SEQ ID NO: 10
 <211> LENGTH: 268
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 10

```

  aacctggcgt tggctcaatgcgt tggctgtttt atgaaatcag accatcagca cagcaatgac
  gatggaaatgt ctaaaatccgt tggatcataca tggatgtatgtt cgtaaatcaaa tcttcacacag
  tggctggatgtt cccatatgtgc tggatgtatgc tggatgtatgc cgtggaaatgg ctgtatgtgc
  ggcctggatgtt atccatgtgc atgtttttgc tggatgtatgc cggacttctgc ctatggccca
  tggtaaaatccaa gtcggaggacgt taaaaggaaac catccatccac atccatcat
  
```

55

56

57

<210> SEQ ID NO: 11
 <211> LENGTH: 96
 <212> TYPE: PRT

-continued

<213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 11

```

    Asn Leu Arg Leu Ser Lys Leu Gly Leu Leu Met Lys Ser Asp His Gln
    1           5           10          15
    His Ser Asn Asp Asp Glu Ser Ser Lys Pro Cys Cys Asp Gln Cys Ala
    20          25          30
    Cys Thr Lys Ser Asn Pro Pro Gln Cys Arg Cys Ser Asp Met Arg Leu
    35          40          45
    Asn Ser Cys His Ser Ala Cys Lys Ser Cys Ile Cys Ala Leu Ser Tyr
    50          55          60
    Pro Ala Gln Cys Phe Cys Val Asp Ile Thr Asp Phe Cys Tyr Glu Pro
    65          70          75          80
    Cys Lys Pro Ser Glu Asp Asp Lys Glu Asn His His His His His His
    85          90          95
  
```

<210> SEQ ID NO 12
 <211> LENGTH: 213
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 12

```

    gacatggaga gctcttaaac ctgttgcgtat caatgcgtat gtaaaggaaatc aatccctcca   60
    cagtggtt gtccgcgtat gggtgtgtat aggtgtgtata gtgcgtatggaa aagctgtatc 120
    tgccgcgtt gtatccgcg tcaatgtttt tgccgtgcaca tcaacggactt ctgtctatgag 180
    ccatggaaac caaaggaggaa cgtatcaaggaa aac                           213
  
```

<210> SEQ ID NO 13
 <211> LENGTH: 71
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 13

```

    Asp Asp Gln Ser Ser Lys Pro Cys Cys Asp Gln Cys Ala Cys Thr Lys
    1           5           10          15
    Ser Asn Pro Pro Gln Cys Arg Cys Ser Asp Met Arg Leu Asn Ser Cys
    20          25          30
    His Ser Ala Cys Lys Ser Cys Ile Cys Ala Leu Ser Tyr Pro Ala Gln
    35          40          45
    Cys Phe Cys Val Asp Ile Thr Asp Phe Cys Tyr Glu Pro Cys Lys Pro
    50          55          60
    Ser Glu Asp Asp Lys Glu Asn
    65          70
  
```

<210> SEQ ID NO 14
 <211> LENGTH: 34
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: primer

<400> SEQUENCE: 14

```

    cagcagccgtt ccagacgtatc agagctctaa accc
  
```

<210> SEQ ID NO 15
 <211> LENGTH: 84
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: primer

<400> SEQUENCE: 15

```

    ctgcagaaac ttaaaaataaa aaaaaacggat ttcccttccgg aaatccgtcc tctgttaact   60
    tttagttctc ttatccgtcc tcgc
  
```

-continued

```

<210> SEQ ID NO 16
<211> LENGTH: 88
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 16
ctgcggaaac ttaaaaataa aaaaacggat ttcccttcagg aaatcccgcc tctgttact
tttaatgggt atgggtatgtc tgggttctc                                         60
86

<210> SEQ ID NO 17
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 17
gatatgcgtc tgaattccgtc tcatatgtca t                                         31
31

<210> SEQ ID NO 18
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 18
atgcactatgc acaggatttc agacgcataat c                                         31
31

<210> SEQ ID NO 19
<211> LENGTH: 74
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 19
cttaaaccgtt ttgcgtatcaa tggcgatgtt ataaatttgta tgggggtggat tggcgatgtca
ggcgatgtcg ttcg                                         60
74

<210> SEQ ID NO 20
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 20
atcttcacgc catatcgctg cagcgacaag tccacccata caaatataaa catgcgttatt
gatgcacaas gggttttagat ct                                         60
92

<210> SEQ ID NO 21
<211> LENGTH: 87
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 21
atttccgttc atatgcgtc caaaagtcgc gcatgttata acctgttacgg gtggacatgt
tttttggat                                         60
67

```

-continued

<211> LENGTH: 67
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 22
 tcgcacgaaa aacaggctca cccgtacagg ttataacatg cgcaagttt gcaggcacts 60
 tgacagg 67

<210> SEQ ID NO 23
 <211> LENGTH: 80
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 23
 ctasaccctg ttgcgcataa tgcgcaatgt tgtttcaggaa ctggggtcac caccgttgtc 60
 gtcgcacgca tgcgtcg 80

<210> SEQ ID NO 24
 <211> LENGTH: 98
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 24
 aatttcacgg cataatcgctg caggcacac ggtgttggacc ccagtccctga acaacacatg 60
 cgcattatc gcaacagggtt tttagact 88

<210> SEQ ID NO 25
 <211> LENGTH: 53
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 25
 caaaagctgt atctgcgttg ttccaggactg gggtcaccac cgttgtttttt gog 53

<210> SEQ ID NO 26
 <211> LENGTH: 61
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 26
 tcgcacgaaa aacaacggtg gtgaccccaag tccctgaacaa cgcagataca gcttttgcatt 60
 g 61

<210> SEQ ID NO 27
 <211> LENGTH: 74
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 27
 cttaaaccctg ttgcgcataa tgcaatgtgtc gtgtaaaaat ccgcatccag tgcgtcg 60
 gcatgtcg tgcg 74

<210> SEQ ID NO 28
 <211> LENGTH: 62
 <212> TYPE: DNA

-continued

```

<210> ORGANISM: Artificial Sequence
<210> FEATURE:
<213> OTHER INFORMATION: synthetic

<400> SEQUENCE: 28
aattcagacg catatcgctg cagcgcacat ggatcggtat tttacgacca cagtcgtcatt      60
gatgcgacca gggttttagag ct                                82

<210> SEQ ID NO 29
<211> LENGTH: 74
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 29
ctaaacccctg ttgcgatcaa tggcggttgct ctcgttctaa cctggacgaa tgtcgtgca      60
gcgatatcgctcgt                                74

<210> SEQ ID NO 30
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 30
aattcagacg catatcgctg cagcgcacatt cgtccagggtt agacgcgacca caaccgcatt      60
gatgcgacca gggttttagag ct                                82

<210> SEQ ID NO 31
<211> LENGTH: 74
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 31
ctaaacccctg ttgcgatcaa tggcggttgct agcgtgtctt gccgatctg tgtcgtgca      60
gcgatatcgctcgt                                74

<210> SEQ ID NO 32
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 32
aattcagacg catatcgctg cagcgcacaca ggatcggtat agcacgtga caaccgcatt      60
gatgcgacca gggttttagag ct                                82

<210> SEQ ID NO 33
<211> LENGTH: 74
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 33
ctaaacccctg ttgcgatcaa tggcggttgct gtcgtctgca catgaaaacc tgtcgtgca      60
gcgatatcgctcgt                                74

<210> SEQ ID NO 34
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

```

-continued

<210> FEATURE:
<213> OTHER INFORMATION: synthetic

<400> SEQUENCE: 34

```
aattcagacg catatcgctg cagcgacagg ttttcatgtg cagacgacca cactggcatt      60
      gatcgacaca gggtttagag ct                                82
```

<210> SEQ ID NO 35
<211> LENGTH: 67
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 35

```
aattcctgcg atatgtgcctg caaaagctgt atctgcgcgc gtagttgcg agctcaatgt      60
      tttttgcg                                67
```

<210> SEQ ID NO 36
<211> LENGTH: 67
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 36

```
tcgacgaaaa aacattgagc tggcacaacta cggggcaga tacagctttt gcaggcacta      60
      tgacacgg                                67
```

<210> SEQ ID NO 37
<211> LENGTH: 74
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 37

```
ctaaacccgt ttgcgcataa tgcaactgtt cgtactcaac ccctccacag tgcgcgtgtca      60
      ggcgcatacg tctg                                74
```

<210> SEQ ID NO 38
<211> LENGTH: 82
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 38

```
aattcagacg catatcgctg cagcgacact gtgggggggt tgagtcgtt cagttgcatt      60
      gatcgacaca gggtttagag ct                                82
```

<210> SEQ ID NO 39
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 39

```
ctgtatctgc aaacgcctaa aatctcgctgg ctgttttgc gtcgcacatca c      51
```

<210> SEQ ID NO 40
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

-continued

```

<400> SEQUENCE: 40
      cgcaaaaaaca gcaacgagat tttgagcgtt tgcagataca gcttttgcat g      51

<210> SEQ ID NO: 41
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 41
      ctgtatctgc tggataatc aaatgacaaac atgtttttgc gtcgacatca c      51

<210> SEQ ID NO: 42
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 42
      cgcaaaaaaca tgggtcatt tgattatacc agcagataca gcttttgcat g      51

<210> SEQ ID NO: 43
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 43
      ctgtatctgc catcaacttg gccccaaattc atgtttttgc gtcgacatca c      51

<210> SEQ ID NO: 44
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 44
      cgcaaaaaaca tggattcggg ccaatgttgc ggcagataca gcttttgcat g      51

<210> SEQ ID NO: 45
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 45
      ctgtatctgc catccgtggg cacccgttcc ttgtttttgc gtcgacatca c      51

<210> SEQ ID NO: 46
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 46
      cgcaaaaaaca agaatacggt gccccacggat ggcagataca gcttttgcat g      51

<210> SEQ ID NO: 47

```

-continued

<211> LENGTH: 51
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

 <400> SEQUENCE: 47

 ctgtatctgc aatcttcatt atcttcasaca gtgttttgc gtgcacatca c 51

<211> SEQ ID NO 48
 <211> LENGTH: 51
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

 <400> SEQUENCE: 48

 cgcaaaaaaca ctgttgaaga taatgaagat tgcagataca gcttttgcat g 51

<211> SEQ ID NO 49
 <211> LENGTH: 51
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

 <400> SEQUENCE: 49

 ctgtatctgc aacccgtctc ttatcgccc gtgttttgc gtgcacatca c 51

<211> SEQ ID NO 50
 <211> LENGTH: 51
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

 <400> SEQUENCE: 50

 cgcaaaaaaca cggggatata aagacgggtg tgcagataca gcttttgcat g 51

<211> SEQ ID NO 51
 <211> LENGTH: 51
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

 <400> SEQUENCE: 51

 ctgtatctgc cttacagatc aatctaaacc gtgttttgc gtgcacatca c 51

<211> SEQ ID NO 52
 <211> LENGTH: 51
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

 <400> SEQUENCE: 52

 cgcaaaaaaca cggtttagat tgatctgtaa ggcagataca gcttttgcat g 51

<211> SEQ ID NO 53
 <211> LENGTH: 51
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

 <400> SEQUENCE: 53

-continued

ctgtatctgc gttacaacat caatgggcat gtgttttc gtcgacatca c	51
<210> SEQ ID NO 54	
<211> LENGTH: 51	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: synthetic oligonucleotide	
<400> SEQUENCE: 54	
cgcaaaaca catggcatt gatgttgtaa cgcagataca gcttttgcat g	51
<210> SEQ ID NO 55	
<211> LENGTH: 51	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: synthetic oligonucleotide	
<400> SEQUENCE: 55	
ctgtatctgc cgccatcac cgtatgtttc gtgttttc gtcgacatca c	51
<210> SEQ ID NO 56	
<211> LENGTH: 51	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: synthetic oligonucleotide	
<400> SEQUENCE: 56	
cgcaaaaca ccaatcatac ggtgtatggc ggcagataca gcttttgcat g	51
<210> SEQ ID NO 57	
<211> LENGTH: 51	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: synthetic oligonucleotide	
<400> SEQUENCE: 57	
ctgtatctgc tcaacacaaa aaattccgca atgttttgc gtcgacatca c	51
<210> SEQ ID NO 58	
<211> LENGTH: 51	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: synthetic oligonucleotide	
<400> SEQUENCE: 58	
cgcaaaaca ttgcggatt ttttgttgtc agcagataca gcttttgcat g	51
<210> SEQ ID NO 59	
<211> LENGTH: 51	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: synthetic oligonucleotide	
<400> SEQUENCE: 59	
ctgtatctgc acacaattc gtcgtcaac atgttttgc gtcgacatca c	51
<210> SEQ ID NO 60	
<211> LENGTH: 51	
<212> TYPE: DNA	

-continued

<210> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide
<400> SEQUENCE: 60
cgcacaaaca tggtcgcagcg cggaaattgtg tgccagataca gcttttgcgt 9 51

<210> SEQ ID NO: 61
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide
<400> SEQUENCE: 61
ctgtatctgc cggcgatcatg ttccgcatct ttgttttgcg gtgcacatca c 51

<210> SEQ ID NO: 62
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide
<400> SEQUENCE: 62
cgcacaaaca aagatggcgaa acatgtatcg ggcagataca gcttttgcgt 9 51

<210> SEQ ID NO: 63
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide
<400> SEQUENCE: 63
ctgtatctgc tcagggtttc cgcttttacatgttttgcg gtgcacatca c 51

<210> SEQ ID NO: 64
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide
<400> SEQUENCE: 64
cgcacaaaca tggtagaaacgc ggaaaggctcg agcagataca gcttttgcgt 9 51

<210> SEQ ID NO: 65
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide
<400> SEQUENCE: 65
tcaatgcgcgt tggaaaggaa tctggactat gctttgcgggt tggcgatc 56

<210> SEQ ID NO: 66
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide
<400> SEQUENCE: 66
cgggacaccggccaaaggatcgatccatgtc cttccatgtcgccatgtatcg ccaagg 57

-continued

```

<210> SEQ ID NO 67
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 67
      caaaagctgt gcttggaaag agatctggac tatgtttgc ttttgcgtcg acatcacgg      59

<210> SEQ ID NO 68
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 68
      aacaaaaaggc aaaggcatagt ccagatcttc tcacaaaggcac agctttgc tgcaactatg      59

<210> SEQ ID NO 69
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 69
      tcaatgcgca tggggggcccttactgtcaaa acatggcggt tgttccgatca tgcgatc      56

<210> SEQ ID NO 70
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 70
      cgaaacacccg gcatgttttg acagtaaggcc cccaaatgcg gcattgtatcg caacagg      57

<210> SEQ ID NO 71
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 71
      caaaagctgt gcttggggcccttactgtcaaaacatgcg ttttgcgtcg acatcacgg      59

<210> SEQ ID NO 72
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 72
      aacaaaaaggc atgttttgac agtaaggcc ccaaaaggcac agctttgc tgcaactatg      59

<210> SEQ ID NO 73
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

```

-continued

<213> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 73

```
tcaatgegcgttgtcttacag tactgtggac tacatgcggg tggccgata tgggtc      56
```

<210> SEQ ID NO: 74

<211> LENGTH: 57

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 74

```
cggaaacccg gcatgtgtgc cacagttactg taagacatgc gcattgtatgc caaacagg      57
```

<210> SEQ ID NO: 75

<211> LENGTH: 59

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 75

```
caaaaatgtgtgttta cagttactgtg gactacatgc ttttgcgtcg acatcacgg      59
```

<210> SEQ ID NO: 76

<211> LENGTH: 59

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 76

```
acgcaaaatgc atgttagtcca cagttactgtta agacaacgc agcttttgc tgcactatg      59
```

<210> SEQ ID NO: 77

<211> LENGTH: 56

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 77

```
tcaatgcgtatgtttt ggaacacgatc tccctggccgg tggccgata tgggtc      56
```

<210> SEQ ID NO: 78

<211> LENGTH: 57

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 78

```
cggaaacccg gcaaggagat ctgttccaaa gagttacatgc gcattgtatgc caaacagg      57
```

<210> SEQ ID NO: 79

<211> LENGTH: 59

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 79

```
cataatgtgtgttgc ttggaaatcg atctccatgc ttttgcgtcg acatcacgg      59
```

-continued

<210> SEQ ID NO 80
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 80
acgcaaaagc aaggagatcg attccaaaaga gtacaaggac agcttttgcg tgcaactatg 59

<210> SEQ ID NO 81
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 81
tcaatgcgca tgtacaaaaca tcgattttac tccttgcggg tgttccgata tgcgtc 59

<210> SEQ ID NO 82
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 82
cggaacacgg gcaaggagta gaatcgatgt ttgtacatgc gcattgtatcg ccaacagg 59

<210> SEQ ID NO 83
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 83
caaaatgtt gtttgccacaa acatcgatcc tactcttgcg ttttggtgtcg acatccacgg 59

<210> SEQ ID NO 84
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 84
acgcaaaaac aaggagatggc atcgatgtt gtgcggacac agcttttgcg tgcaactatg 59

<210> SEQ ID NO 85
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 85
tcaatgcgca tgtacaaaaaa tcgatcgatc tccttgcggg tgttccgata tgcgtc 59

<210> SEQ ID NO 86
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

-continued

```

<400> SEQUENCE: 86
      cgaaacacgg gcaaggagta cgtcgattt ttgtacatgc gcattgtatc caacagg      57

<210> SEQ ID NO: 87
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 87
      caaaagctgt gtttgcacaa aatcgatcg tactccgttgt ttttgcgtcg acatcacgg      59

<210> SEQ ID NO: 88
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 88
      acgcacaaac aaggagtcg atcgattttt gtgcacgcac agcttttgc a tgcactatg      59

<210> SEQ ID NO: 89
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 89
      tcaatgcgcg tgcacccgc agacaaactga aacatgcggg tgcacccata tgcgtc      56

<210> SEQ ID NO: 90
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 90
      cggaacacgg cgtatgttca gttgtctgca gggtacatgc gcattgtatc caacagg      57

<210> SEQ ID NO: 91
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 91
      caaaagctgt gtttgcaccc tgcacacaaac tgcacatgt ttttgcgtcg acatcacgg      59

<210> SEQ ID NO: 92
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 92
      acgcacaaac atgtttcgt tgcacccagg tggcaacgcac agctttgc a tgcactatg      59

<210> SEQ ID NO: 93
<211> LENGTH: 56

```

-continued

-continued

-continued

```

<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 106
      acgcaaaaaac agttaagagt aggtgttaac aggccaggcac agcttttgc a tgcactatg      59

<210> SEQ ID NO 107
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 107
      tcaatgcgca tgcgttcc caacttacca taaactgttgg tggttccgata tggtgtt      57

<210> SEQ ID NO 108
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 108
      cggaaacacog acgatgttggaa tgagttggaa gagccatgc gcattgtatc caacagg      57

<210> SEQ ID NO 109
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 109
      caaaatgtt gctgtcgcc ttctacacaca ctctactgt ttttgcgtcg acatcaagg      59

<210> SEQ ID NO 110
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 110
      aacgcaaaaaac agttagatgt tgtaggaaagc ggcaggcac agcttttgc a tgcactatg      59

<210> SEQ ID NO 111
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 111
      caaaatgtt gctgtcgcc ttggccatttgc cccacattgt ttttgcgtcg acatcaagg      59

<210> SEQ ID NO 112
<211> LENGTH: 58
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 112
      acgcaaaaaac aagggtggca aaggccataaa gggccaggcac agcttttgc a tgcactat      58

```

-continued

```

<210> SEQ ID NO 113
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 113
      aagctgtatac tgcttggaca tcgattctac accttggttt tgcgctgaca tcacgg      56

<210> SEQ ID NO 114
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 114
      acgcacaaac aagggtgtaca atcgatgttc cagcagatac agcttttgc tgcact      56

<210> SEQ ID NO 115
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 115
      gcgatcaatg catctgtact tggatttgaca gtaacttcttgc tgggtgttcc gatatgggtc      60

<210> SEQ ID NO 116
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 116
      ggaacacccga caagggtgtac tgtcaatccca agtacagatgc cattgtatgc aacagggtt      60

<210> SEQ ID NO 117
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 117
      aagctgtatac tgacatggaa tcgatgttac tccttggttt tgcgctgaca tcacgg      56

<210> SEQ ID NO 118
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 118
      acgcacaaac aagggtgtaca atcgatccat gtgcagatac agcttttgc tgcact      56

<210> SEQ ID NO 119
<211> LENGTH: 56
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

```

-continued

<400> SEQUENCE: 119
 aaggtgtata tgtacatgga tggatggac accttggttt tgctcgaca tcaagg 56

<210> SEQ ID NO 120
 <211> LENGTH: 56
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 120
 acgcaaaaaac aagggtgtca atogatocat gtacagatac agcttttgc tgcact 56

<210> SEQ ID NO 121
 <211> LENGTH: 59
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 121
 caaaagctgc gcatgtgtta ctacagattt gatcgaatgt ttttgcgtcg acatcaagg 59

<210> SEQ ID NO 122
 <211> LENGTH: 59
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 122
 acgcaaaaaac attcgatcca atctgtgtta acacatgcgc agcttttgc tgcaactatg 59

<210> SEQ ID NO 123
 <211> LENGTH: 61
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 123
 caaaaggctgt gctgtgccaa cactttggac tcataatgtt ttttgcgtcg acatcaagg 60
 c 61

<210> SEQ ID NO 124
 <211> LENGTH: 61
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 124
 acgcaaaaaac acatatgtgtt ccaaaggctgtt gggcaggcgc agcttttgc tgcaactatg 60
 c 61

<210> SEQ ID NO 125
 <211> LENGTH: 59
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 125
 caaaaggctgc gcatgtttact acttcattt ccaccaatgt ttttgcgtcg acatcaagg 59

-continued

```

<210> SEQ ID NO 126
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE 126
acgcaaaaac atgggtggaa ttgagatgt taacatgcgc agcttttgc tgcactatg      59

<210> SEQ ID NO 127
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE 127
caaaagctgt ctttgtccgg aaaaacgtttaa cgttttctt tggtaatttgcg tggacatcac      60
ggacttctgg                                         69

<210> SEQ ID NO 128
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE 128
tgtcgacca attacaaggaa gaaaacgtttaa cgttttccgg aaaaaggacag cttttgcgtg      60
cactatggac                                         69

<210> SEQ ID NO 129
<211> LENGTH: 47
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE 129
caaaaacgtgt gctttgtaaaaac aaaaacgttacg tctttttatgt tttttgcg      47

<210> SEQ ID NO 130
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE 130
tcgacgcacaa aacataaasag acgttacgttgc tgtttacaag cacagtttgcgtg      55

<210> SEQ ID NO 131
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE 131
gatcccaaggat gacgtgtttt gtttgcgtt gatgtgtt      37

<210> SEQ ID NO 132
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

```

-continued

<210> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 132

```
ctccatcgatca actaaaggcag ctccacactg
```

29

<210> SEQ ID NO: 133
<211> LENGTH: 70
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 133

```
gatccagggt acacctgaccc aacctcccca tctgtatctg gagaataccg agcttgggac  
gatggagact
```

60
70

<210> SEQ ID NO: 134
<211> LENGTH: 62
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 134

```
ctccatcgatcc caaaggctgggt attctccagg atcagatgga ggtagtgggt cagggttacc  
tg
```

60
62

<210> SEQ ID NO: 135
<211> LENGTH: 79
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 135

```
gatcccgccg acctcgccgt gtctaaaggctt ggccgttctt tggaaatcaga ccatacgac  
ggcaatgacg atgagagat
```

60
79

<210> SEQ ID NO: 136
<211> LENGTH: 71
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 136

```
ctccatcgatca ttgtctgtct gatggctctga tttcataaaggc agggcaagct tagacacgacg  
cagggttccgg
```

60
71

<210> SEQ ID NO: 137
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 137

```
gatcccaat cggccatca gcaacgaaat gacgtatgaga gct
```

43

<210> SEQ ID NO: 138
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

-continued

<400> SEQUENCE: 138

ctcatcgtca ttgtgtgtct gatggctctga ttttg

35

<210> SEQ ID NO: 139

<211> LENGTH: 205

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 139

gtatccaaatgg accccggaccc	aaacggccccc	aaatgtatcc	ggatgtggat	60	
tcataatccaa	tttacacaaa	tggaaatgtgt	tatccataacg	gtcagtatgc	120
tgttgtgtat	aaaaatcaaga	ggccatgtac	ccatatcgctc	cgtggggacc	180
gatccaggat	cacatgtatc	gagct	actccaaatct	205	

<210> SEQ ID NO: 140

<211> LENGTH: 197

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 140

ctcatcgtct	gtatctgggt	cagatgttgg	tgtttccca	ggacgtatgc	60
ctcttgattt	tgtgtccacc	attttgtgt	ccataactga	cgtttagatgc	120
attttgtgtat	attttgtgtt	aaaaatccatgc	tggatctct	cgtggatcac	180
tgggtcggt	tctccgt		ttgggggggt		197

<210> SEQ ID NO: 141

<211> LENGTH: 6

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: linker peptide

<400> SEQUENCE: 141

Trp Gly Asp Pro His Tyr
1 5

<210> SEQ ID NO: 142

<211> LENGTH: 6

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: linker peptide

<400> SEQUENCE: 142

Asp Asn Asn Asp Pro Ile
1 5

<210> SEQ ID NO: 143

<211> LENGTH: 6

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: linker peptide

<400> SEQUENCE: 143

Val Val Ala Asp Pro Asn
1 5

<210> SEQ ID NO: 144

<211> LENGTH: 37

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

-continued

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 144

```
tggcggtca g caacatgagc ggcgggtcg atgat 37
```

<210> SEQ ID NO: 145

<211> LENGTH: 37

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 145

```
taatcatcg ctggcgctg catgttgtg aacgc 37
```

<210> SEQ ID NO: 146

<211> LENGTH: 128

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 146

```
gacatcaagg acttctgtca tgagccatgt aaaccaaggc aggacgtaa agagaactaa 60
aaaccttaact cgaggtaac agggacggca ttccctggaa gaaatccgtt tttttttt 120
taattaaag 128
```

<210> SEQ ID NO: 147

<211> LENGTH: 110

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 147

```
agctcttaat taaaataaaa aaaaacggatt ttccctggaa aatccgtctt ctgtttaacct 60
cgatgttaac tttagttctt ctttatcgto ctcgttgtt ttatcggtt catggcggaa 120
gtccgtgtat 130
```

<210> SEQ ID NO: 148

<211> LENGTH: 23

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 148

```
caccaacatg a g c g c g c a g g 23
```

<210> SEQ ID NO: 149

<211> LENGTH: 59

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 149

```
atcgcttgg a tccggatagt gggggctcc ccaagatgtt gattcttta tttttccc 59
```

<210> SEQ ID NO: 150

<211> LENGTH: 59

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 150

-continued

```

atcgctgga tcgggtatgg gatcattgtt gtcaagatgt gattcttcta ttttttccc      59

<210> SEQ ID NO 151
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 151
atcgctgga tcgggtatgg gatctgcaac tacagatgtt gattcttcta ttttttccc      59

<210> SEQ ID NO 152
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 152
gcataaggat gatgtcatctg cagcg      25

<210> SEQ ID NO 153
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 153
atcgctgga tcggatagt gggggctcc ccacggttct cttggatcg atggggg      57

<210> SEQ ID NO 154
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 154
atcgctgga tcggatatgg gatcattgtt gtcgggttct cttggatcg atggggg      57

<210> SEQ ID NO 155
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 155
atcgctgga tcgggtatgg gatctgcaac tacgggttct cttggatcg atggggg      57

<210> SEQ ID NO 156
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: linker peptide

<400> SEQUENCE: 156
Trp Gly Arg Pro His Tyr
  1           5

<210> SEQ ID NO 157
<211> LENGTH: 6

```

-continued

```

<210> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: linker peptide

<400> SEQUENCE: 157
Asp Asn Asn Asp Pro Ile
1 5

<210> SEQ ID NO 158
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: linker peptide

<400> SEQUENCE: 158
Val Val Ala Asp Pro Asn
1 5

<210> SEQ ID NO 159
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: linker peptide

<400> SEQUENCE: 159
Ile Pro Pro Ser Asp Pro Thr Pro Pro Ser Asp Pro Gly Glu Pro Trp
1 5 10 15
Gly Asp Pro His Tyr
20

<210> SEQ ID NO 160
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: linker peptide

<400> SEQUENCE: 160
Ile Pro Pro Ser Asp Pro Thr Pro Pro Ser Asp Pro Gly Glu Pro Asp
1 5 10 15
Asn Asn Asp Pro Ile
20

<210> SEQ ID NO 161
<211> LENGTH: 21
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: linker peptide

<400> SEQUENCE: 161
Ile Pro Pro Ser Asp Pro Thr Pro Pro Ser Asp Pro Gly Glu Pro Val
1 5 10 15
Val Ala Asp Pro Asn
20

<210> SEQ ID NO 162
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 162
gateccagggtg gggacgacgca tgacaaaggac gatggaggt

```

-continued

```

<210> SEQ ID NO 163
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 163
ctcatcgtct ttgtcatacgt cgtctccacc tg                                32

<210> SEQ ID NO 164
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 164
gatccagggtg ctgttcattt caaacatgtat agtt                                34

<210> SEQ ID NO 165
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 165
ctcatcgtcg taatgagcgtg caccctg                                26

<210> SEQ ID NO 166
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 166
gatccacgtg ctaaaagaga caatgtttttt t                                31

<210> SEQ ID NO 167
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 167
ctcatcgtct ttttttagcac gtg                                23

<210> SEQ ID NO 168
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 168
gatccaggcg ctgcacacata caaacatgtat gagatgt                                37

<210> SEQ ID NO 169
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

```

-continued

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 169

ctcatcgctcg ttgtatgtgtc cagcgccctg

29

<210> SEQ ID NO 170

<211> LENGTH: 28

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 170

gatccattcc ttgaaggacga tgagagct

28

<210> SEQ ID NO 171

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 171

ctcatcgctc tcaaggaatg

20

<210> SEQ ID NO 172

<211> LENGTH: 29

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 172

cccatacccgg agccagacga tgagagctc

29

<210> SEQ ID NO 173

<211> LENGTH: 32

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 173

catcgctcg ctccggatgtt ggatcatttg tg

32

<210> SEQ ID NO 174

<211> LENGTH: 21

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: linker peptide

<400> SEQUENCE: 174

Asp Asn Asn Asp Pro Ile Pro Glu Pro Asp Asp Glu Ser Phe Asn Met
1 5 10 15
Pro Ile Pro Glu Pro

20

<210> SEQ ID NO 175

<211> LENGTH: 58

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic oligonucleotide

<400> SEQUENCE: 175

-continued

 gatccaggcg ctgcacacta caaatcggac catcgacca gcaatgacga tgagact 58

<210> SEQ ID NO 176
 <211> LENGTH: 50
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide
 <400> SEQUENCE: 176
 ctcatcgta ttgcgtgtctgatgggtctgtttgtgtgtgcagcgcctg 50

<210> SEQ ID NO 177
 <211> LENGTH: 46
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide
 <400> SEQUENCE: 177
 gatccaggcg ctgcacacta cgtagaattt caagacgtg agagct 46

<210> SEQ ID NO 178
 <211> LENGTH: 38
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic oligonucleotide
 <400> SEQUENCE: 178
 ctcatcgct tgaaattcta cgtgtgtgc agcgcctg 38

<210> SEQ ID NO 179
 <211> LENGTH: 18
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: cleavage site
 <400> SEQUENCE: 179
 Asp Asn Asn Asp Pro Ile Pro Asp Pro Gly Ala Ala His Tyr Val Glu
 1 5 10 15
 Phe Glu
 <210> SEQ ID NO 180
 <211> LENGTH: 87
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic fusion peptide
 <400> SEQUENCE: 180

Lys	Ile	Arg	Gln	Ser	Ala	Ser	Asp	Asn	Asn	Asp	Pro	Ile	Pro	Asp	Pro
1	5	9	10	11	15	19	20	24	25	26	27	30	31	35	36
Asp	Asp	Glu	Ser	Lys	Pro	Cys	Cys	Asp	Gln	Cys	Ala	Cys	Thr	Lys	Asn
20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Ser	Asn	Pro	Pro	Gln	Cys	Arg	Cys	Ser	Asp	Met	Arg	Leu	Asn	Ser	Cys
35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
His	Ser	Ala	Cys	Lys	Ser	Cys	Ala	Cys	Tyr	Asn	Leu	Tyr	Gly	Trp	Thr
50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65
Cys	Phe	Cys	Val	Asp	Ile	Thr	Asp	Phe	Cys	Tyr	Glu	Pro	Cys	Lys	Pro
65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Ser	Glu	Asp	Asp	Lys	Glu	Asn									
85															

<210> SEQ ID NO 181
 <211> LENGTH: 46
 <212> TYPE: DNA

-continued

```

<210> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide primer

<400> SEQUENCE: 181
aacatgacgcg cgcaggctgt tgacgcggca attcaacaaa cgttag 46

<210> SEQ ID NO: 182
<211> LENGTH: 77
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide primer

<400> SEQUENCE: 182
tgcgttggat cccgttatggg atcattgttg tcaccagaac cactagtttg tcccttaccg 60
ctggatattt tttgggt 77

<210> SEQ ID NO: 183
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide primer

<400> SEQUENCE: 183
tgcacttc tgcgtttggtc tggtaacgc agaggcttgc acaacaatgtt tccattatccg 60

<210> SEQ ID NO: 184
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic oligonucleotide primer

<400> SEQUENCE: 184
gatccggat aggatcattt ttgtcaagac ctctgcgttc aacagacccaa agcggaggaa 60

```

40

We claim:

1. A method for producing a VEGF binding peptide in a protease inhibitor scaffold in a bacterial cell comprising, introducing a DNA construct into a bacterial cell, wherein said DNA construct comprises a heterologous DNA sequence encoding a protease inhibitor, said inhibitor being a Bowman-Birk Inhibitor (BBI) comprising SEQ ID NO: 13, wherein the trypsin and/or chymotrypsin loop of SEQ ID NO: 13 is replaced with a VEGF binding peptide consisting of SEQ ID NO: 9; culturing said bacterial cell under suitable culture conditions to allow expression of the heterologous DNA sequence, and producing said VEGF binding peptide in a protease inhibitor scaffold.
2. The method of claim 1, further comprising recovering said VEGF binding peptide in a protease inhibitor scaffold.
3. The method of claim 2, further comprising the step of activating said VEGF binding peptide in a protease inhibitor scaffold.
4. The method of claim 3, wherein said activating is accomplished by exposing said VEGF binding peptide in a protease inhibitor scaffold to at least one reagent selected from the group consisting of compositions that reduce disulfide bonds, compositions that oxidize disulfide bonds, and compositions that alter the redox potential.

5. The method of claim 1, wherein said bacterial cell is a member of the genus *Bacillus*.

6. The method of claim 1, further comprising introducing into the bacterial cell a second nucleic acid sequence encoding a thiol disulfide oxidoreductase or chaperone.

7. The method of claim 1, wherein said VEGF binding peptide in a protease inhibitor scaffold is expressed as a fusion protein.

8. The method of claim 7, wherein said fusion protein further comprises a cellulase catalytic domain, a cleavage site, and said VEGF binding peptide in a protease inhibitor scaffold.

9. The method of claim 7, wherein said fusion protein is processed by a protease or acid/heat treatment to liberate said VEGF binding peptide in a protease inhibitor scaffold.

10. The method of claim 7, wherein the fusion protein further comprises at least one linker sequence.

11. The method of claim 10, wherein said linker sequence is selected from the group consisting of SEQ ID NOS: 141-143.

12. An isolated polynucleotide encoding a Bowman-Birk Inhibitor (BBI) comprising SEQ ID NO: 13, wherein the trypsin and/or chymotrypsin loop of SEQ ID NO: 13 is replaced with SEQ ID NO: 9.

13. An expression vector comprising a polynucleotide sequence, wherein said polynucleotide sequence encodes a

45

55

55

55

60

65

113

Bowman-Birk inhibitor comprising SEQ ID NO: 13, wherein the trypsin and/or chymotrypsin loop of SEQ ID NO: 13 is replaced with SEQ ID NO: 9.

14. A host cell transformed with the vector of claim 13.

15. The host cell of claim 14, wherein said host cell is a ⁵ *Bacillus* species cell.

16. A method for producing a VEGF binding peptide in a protease inhibitor scaffold in a bacterial cell comprising,

introducing a DNA construct into a bacterial cell, wherein said DNA construct comprises a heterologous DNA ¹⁰ sequence encoding a protease inhibitor, said inhibitor

being a Bowman-Birk Inhibitor (BBI) comprising SEQ ID NO: 13, wherein the trypsin and/or chymotrypsin loop

114

of SEQ ID NO: 13 is replaced with a VEGF binding peptide consisting of SEQ ID NO: 9, and wherein SEQ ID NO: 13 is truncated at the N- and/or C-terminus, with as many as seven N-terminal residues up to the first cysteine (Cys8) and/or as many as nine C-terminal residues up to the last cysteine (Cys62) removed; culturing said bacterial cell under suitable culture conditions to allow expression of the heterologous DNA sequence, and producing said VEGF binding peptide in a protease inhibitor scaffold.

* * * * *